
ANÀLISI I INTEGRACIÓ A UNA BASE DE DADES RELACIONAL DELS REGISTRES DE BAPTISMES D'UNA POBLACIÓ CATALANA.

Memòria del Treball Final de Grau

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona



Autor: Nicola Bafundi Angueira

Director: Enric Mayol

Grau en Enginyeria Informàtica

Especialitat de Sistemes d'Informació

Resum

Aquest projecte té com a finalitat la creació d'una eina que faciliti la realització d'anàlisis i estudis genealògics. Per a la seva elaboració es disposa d'un conjunt de fitxers Excel amb registres dels baptismes d'una població catalana, en la qual és difícil realitzar la cerca d'avantpassats.

En primer lloc, es dissenyarà i crearà una base de dades relacional que, mitjançant un procés d'extracció, transformació i càrrega, emmagatzemarà les dades provinents dels registres. D'aquesta manera, s'aconsegueix la possibilitat de fer consultes SQL, amb tot el potencial que això comporta.

En segon lloc, serà necessària la creació d'una aplicació web connectada a la base de dades que serveixi com a plataforma per accedir a les dades i que permeti la realització de diversos tipus de consultes, prèviament dissenyades i parametrizables, de les quals s'obtinguin resultats interessants pels estudis genealògics.

En conclusió, es crearà una eina informàtica que permetrà donar resposta a les cerques d'avantpassats amb rapidesa i comoditat.

Resumen

Este proyecto tiene la finalidad de crear una herramienta que facilite la realización de análisis y estudios genealógicos. Para su elaboración se dispone de un conjunto de ficheros Excel con los registros de bautizos de una población catalana, en la cual es particularmente difícil realizar la búsqueda de antepasados.

En primer lugar, se diseñará y creará una base de datos relacional que, mediante un proceso de extracción, transformación y carga, almacenará los datos provenientes de los registros. De esta manera, se consigue la posibilidad de hacer consultas SQL, con todo el potencial que eso conlleva.

En segundo lugar, será necesaria la creación de una aplicación web conectada a la base de datos que sirva como plataforma para acceder a los datos y permita la realización de diferentes tipos de consultas, previamente diseñadas y parametrizables, de las cuales se obtengan resultados interesantes para los estudios genealógicos.

En conclusión, se creará una herramienta informática que permitirá dar respuesta a las búsquedas de antepasados con comodidad y velocidad.

Abstract

This project has the purpose of creating a tool that facilitates the analysis and genealogical studies. To create it, there is a set of Excel files with the records of baptisms of a Catalan population, where it is particularly difficult to search for ancestors.

First, a relational database will be designed and created to store the data coming from the registers through a process of extraction, transformation and loading. By this way, you get the possibility of doing SQL (Structured Query Language) queries, with all the potential that entails.

Second, it will be necessary to create a web application, connected to the database, that serves as a platform to access the data and allows the realization of different types of queries, previously designed and parameterized, from which interesting results for the genealogical studies can be obtained.

In conclusion, a computer tool will be created and will allow the search of ancestors with comfort and speed.

Agraïments

Primer de tot, agraeixo al professor Enric Mayol per la seva supervisió, aportació d'idees i consells, durant la realització d'aquest treball.

També vull agrair a les persones que van fer la tasca de migrar les dades dels arxius originals dels registres als fitxers Excel utilitzats, ja que sense ells no hagués estat possible la realització d'aquest projecte.

Finalment, vull agrair a la meva família, amics i companys pel suport i la confiança que m'han donat durant aquesta etapa dels meus estudis universitaris.

Índex

1. Introducció i Estat de l'art.....	6
1.1 Context.....	6
1.1.1. Actors implicats.....	7
1.2. Formulació del problema	8
1.3. Estat de l'art.....	9
1.3.1. Projecte “Enxarxeu-ho!” de la SCGHSVN	9
1.3.2. Five centuries of Marriages (Cinc segles de Matrimonis)	10
1.3.3. Family Search	11
1.3.4. Conclusió de l'estat de l'art.....	12
2. Abast del projecte.....	13
2.1 Obstacles.....	14
2.2 Metodologia de treball.....	15
2.3 Eines de seguiment i anàlisi.	16
3. Planificació temporal.....	17
3.1 Descripció de tasques.....	17
3.2 Diagrama de Gantt	18
3.3 Distribució del temps i Recursos.....	19
3.3.1 Recursos humans.....	19
3.3.2 Recursos materials.....	19
3.4 Planificació d'alternatives i pla d'acció.....	20
4. Pressupost i Sostenibilitat	21
4.1. Autoavaluació del domini actual de la competència de sostenibilitat	21
4.2. Dimensió Econòmica	22
4.2.1 Identificació i estimació dels costos	22
4.2.2 Control de Gestió.....	23
4.2.3 Reflexió.....	24
4.3 Dimensió Ambiental	25
4.4 Dimensió Social.....	25
5. Definició de requisits.....	26

5.1. Requisits Funcionals	26
5.2. Requisits de qualitat	26
5.3. Restriccions	27
6. Disseny de la base de dades	27
6.1. Disseny conceptual en UML.....	27
6.2. Disseny de les taules de la BBDD	32
6.3. Elecció del Sistema Gestor de Base de Dades.....	34
7. Procés Extraction-Transform-Load (ETL).....	35
7.1. Mètode seguit.....	35
7.2. Dificultats trobades i condicions a seguir en el procés ETL.....	38
8. Disseny i Implementació de l'aplicació web	41
8.1. Framework escollit	41
8.2. Primers passos de l'aplicació web	42
8.3. Creació i disseny d'una pàgina web.....	45
8.4. Consultes implementades.....	47
9. Resultats.....	56
9.1. Objectius assolits	56
9.2. Futur del projecte	56
9.3. Aportacions del projecte	57
10. Annexos.....	58
11. Bibliografia	59

1. Introducció i Estat de l'art

Aquest projecte és un Treball Final de Grau emmarcat dins l'especialitat de Sistemes d'Informació del Grau d'Enginyeria Informàtica impartit a la Facultat d'Informàtica de Barcelona de la Universitat Politècnica de Catalunya. Consisteix en la migració dels registres de baptismes d'una població catalana d'uns fitxers Excel a una base de dades relacional per al seu posterior anàlisi des d'una pàgina web.

1.1 Context.

L'anàlisi de registres de baptismes és rellevant per la realització d'estudis genealògics. La genealogia és la disciplina que es dedica a estudiar l'origen i la successió de les persones, en concret de les famílies. Pot complementar altres ciències, com la medicina, el dret, la sociologia, la nobiliària, la genètica, l'estadística, l'economia, la paleontologia... i és de necessitat en casos com afers d'herències o en medicina patològica per poder detectar les malalties congènites.

A Catalunya, es va fundar l'any 1983 la Societat Catalana de Genealogia, Heràldica, Sigil·lografia, Vexil·lologia i Nobiliària (SCGHSVN) amb l'objectiu de fer conèixer aquestes disciplines mitjançant diferents seminaris i exposicions.



1. Logotip de la Societat Catalana de Genealogia

Una de les tasques que realitzen els seus membres, es la de recol·lectar informació d'arxius històrics i registres municipals o parroquials on es feia el cens de baptismes, matrimonis i/o obituàries i fer un "buidatge", és a dir, extraure les dades bàsiques de cadascun dels registres en un fitxer Excel, en el qual cada bateig o matrimoni ocupi una fila. Aquesta tasca es pot realitzar o bé a mà, o bé mitjançant tècniques de reconeixement òptic de caràcters (Optical Character Recognition, OCR).

Un d'aquestes buidatges és el que es tractarà en aquest projecte, en concret el de la població catalana de nom Agullana, situada al nord de Catalunya, a la província de Girona. [7]. El buidatge consta de 31 fitxers Excel per a baptismes, 5 per a matrimonis i 8 per a obituàries, cadascun amb un contingut que oscil·la entre 300 i 800 registres per fitxer.



Figura 2. Església de Santa Maria d'Agullana. Font: [2]

1.1.1. Actors implicats

Durant la realització d'aquest projecte i un cop estigui acabat, hi haurà certs actors involucrats en les seves fases.

- Director del projecte: El director és l'Enric Mayol, qui s'encarregarà de supervisar i guiar el projecte.
- Proveïdors dels registres: Els encarregats de fer el buidatge de l'arxiu físic als fitxers Excel en un format tractable.
- Desenvolupadors: L'equip de desenvolupament està format per diversos rols: l'enginyer de requisits, el dissenyador de la base de dades i del procés ETL, l'administrador de la base de dades relacional i el desenvolupador Front-End per a la pàgina web. En aquest cas, tots aquests rols seran adoptats per l'autor del projecte.
- Usuaris finals: Totes aquelles persones interessades en fer diferents consultes i anàlisis relacionats amb els registres de baptismes d'Agullana, en especial els genealogistes i els mateixos habitants d'Agullana i les seves proximitats.

1.2. Formulació del problema

Les eines de suport a les recerques genealògiques no estan del tot desenvolupades a Catalunya i al Estat Espanyol, amb un clar endarreriment respecte altres països. Per aquest motiu es troba la necessitat de crear una nova eina, on no només es pugui fer recerca a registres per noms i cognoms, si no també on es puguin fer diferents tipus de consultes més complexes que responguin a la gran quantitat de feina que suposa el trasllat de registres de cens demogràfic a fitxers informàtics.

Per a aquest projecte es disposa del buidatge de registres de baptismes, matrimonis i obituaris de la parròquia de la població catalana d'Agullana, en una considerable quantitat de fitxers Excel on es fa difícil la cerca dels avantpassats i d'establir relacions entre les persones que estan llistades.

Donat això, aquest projecte s'estableix dos objectius principals:

- **Creació d'una base de dades relacional** on es traslladi l'informació d'aquests fitxers Excel i on el seu disseny doni la possibilitat de fer diverses consultes SQL interessants per estudis genealògics.
- **Construcció d'una pàgina web** que aporti accessibilitat a la base de dades creada, que permeti la realització d'aquestes consultes a usuaris no especialitzats i que aporti estadístiques i anàlisis d'interès general sobre les dades.

Un cop finalitzat, es podria estendre al projecte "Enxarxeu-ho!" de la SCGHSVN, del qual es parlarà més endavant, i podria servir d'exemple per a motivar l'aparició d'altres projectes de buidatges i d'explotació d'aquestes dades, demostrant que es pot treure més profit de la dura tasca dels processos de buidatges.

1.3. Estat de l'art

Aquest projecte té com objectiu la creació d'una pàgina web on es puguin consultar diversos registres d'una població per trobar avantpassats i per motius genealògics en general, però ja existeixen altres pàgines o projectes que tenen un objectiu similar.

1.3.1. Projecte "Enxarxeu-ho!" de la SCGHSVN

La Societat Catalana de Genealogia ha iniciat aquest projecte pilot amb l'objectiu de poder donar accés als seus socis i a tots els interessats en les recerques genealògiques a un dipòsit gran de fitxers de buidatges de llibres sacramentals i semblants disponibles a Internet.

Volen publicar, en una pàgina web, un inventari de tots els materials ja disponibles a Internet referents a localitats catalanes i també un inventari amb els buidatges elaborats de forma manual i que és podrien convertir en fitxers informàtics ja sigui a mà o amb tècniques de visió per computador. Començarien per aquells amb més demanda i amb més potencial de ser utilitzats per futurs treballs de recerca, seguits per aquells on els arxivadors estiguin més predisposats a participar en el projecte i conviden a tots els interessats que vulguin aportar treball i informació d'una determinada localitat.

També tenen com a objectiu realitzar propostes de models de treballs que motivin la participació de voluntaris en la elaboració d'aquests buidatges, garantint la qualitat.

Mitjançant aquest projecte, volen omplir el buit de la manca d'informació i organització que hi ha a Catalunya i en general a tota la Península Ibèrica en l'àmbit de la genealogia. Disposar de tota aquesta informació a Internet permetria fer recerca genealògica sense haver-se de desplaçar a la localitat en concret on es troben els llibres de registre, estalviaria als arxivers dedicació per atendre el públic i estalviaria als mateixos llibres manipulacions no necessàries que empitjorarien el seu estat. A més, el fet de facilitar l'accés a aquest tipus de dades probablement incrementarà el nombre d'interessats en la genealogia. [5]

1.3.2. Five centuries of Marriages (Cinc segles de Matrimonis)

Aquest projecte és una iniciativa de recerca a llarg termini, basat en la explotació de les dades dels “Llibres de les Esposalles” conservats a l’Arxiu de la Catedral de Barcelona, una font de dades formada per 244 llibres de registres de matrimonis amb informació sobre més de mig milió de casaments celebrats entre el 1451 i 1905 en aproximadament 250 parròquies.

Aquests llibres permeten la realització de recerques demogràfiques de gran recorregut temporal i de gran extensió territorial, des del centre de la ciutat fins als pobles rurals del seu voltant.

Es poden fer recerques relacionades amb estimacions de població, estimacions indirectes de fertilitat, migració i supervivència, així com altres estudis socioeconòmics com l’evolució de les classes socials en funció de les taxes aplicades als matrimonis.

Això ja es pot fer amb més facilitat gràcies a l’assoliment de l’objectiu d’aquest projecte, que és la transformació del “Llibre de les Esposalles” en una biblioteca digital, afavorint la conservació d’aquest document històric, que ha sobreviscut al llarg de cinc segles a revoltes i destruccions de documents eclesiàstics.

Aquesta biblioteca digital és una aplicació web realitzada amb la col·laboració de la Universitat Autònoma de Barcelona, el Centre d’Estudis Demogràfics i el Centre de Visió per Computador, aquest últim ajudant al reconeixement de caràcters per part del ordinador per accelerar la digitalització dels manuscrits. [1][3]

Aquesta és la pàgina inicial de la web: <http://dagapp.cvc.uab.es/5cofm/>



Figura 3. Captura de pantalla de la pàgina inicial de Five Centuries of Marriages

Malauradament només poden entrar usuaris autoritzats i per fer el registre s’ha de sol·licitar aprovació.

1.3.3. Family Search

Family search és una organització sense ànim de lucre amb l'objectiu principal d'ajudar a les persones a trobar els seus avantpassats donant accessibilitat online a registres històrics. És l'organització genealògica més gran del món, posicionant-se a la capdavantera del sector en l'àmbit de recopilació, indexació i conservació de registres.

Els seus serveis són gratuïts, tenen centres repartits globalment i la seva base de dades està compresa per més de 4 mil milions de noms de tot el món.

La seva pàgina web permet utilitzar els seus serveis gratuïtament mitjançant el registre. Un cop fet, permet realitzar l'arbre genealògic de l'usuari demanant-li dades sobre els seus pares, avis i besavis i dóna eines per fer la ressenya bibliogràfica. Per tal d'omplir els buits de l'arbre, també et proporciona un cercador que busca entre la seva base de dades de registres i en altres arbres genealògics d'usuaris inscrits, sempre i quan aquests ho hagin fet públic.

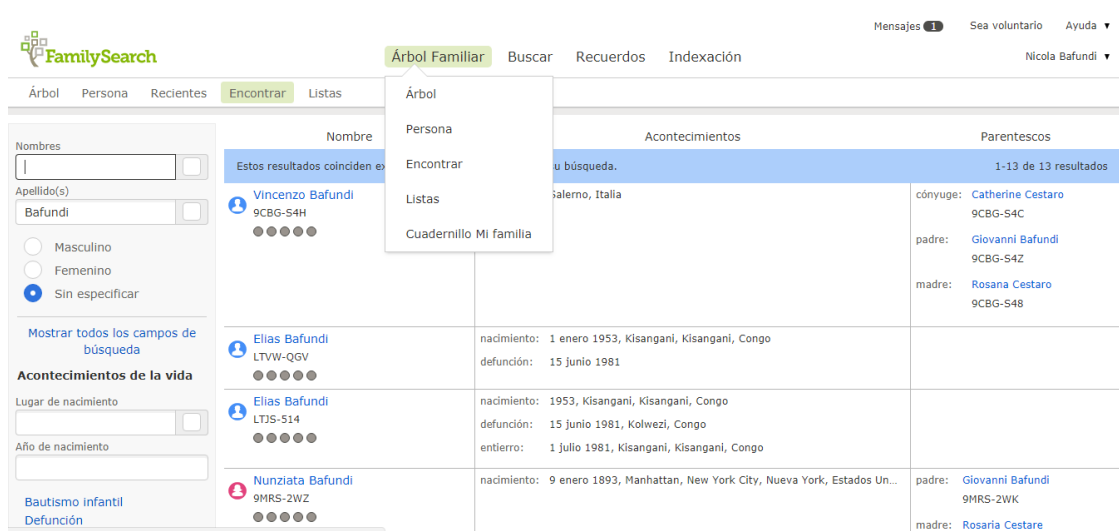


Figura 4. Captura de pantalla de Family Search

Com es pot veure a la captura anterior, es pot buscar pel nom i cognoms de les persones, filtrar per sexe, per lloc i any de naixement, si està a registres de defuncions, baptismes o matrimonis, etc.

Als resultats de la cerca, també proporciona informació sobre qui van ser els pares i la parella, donant profunditat a la cerca per ampliar les branques de l'arbre genealògic, si es disposa de l'informació.

Una de les funcionalitats a destacar és la cerca d'avantpassats en registres de tot el món, tant de naixements, matrimonis, defuncions o de residència i amb la possibilitat d'utilitzar una gran varietat de filtres per acotar la cerca i obtenir els resultats que es desitgin.

2. Abast del projecte

La finalitat d'aquest projecte és la creació d'un sistema que constarà de les següents parts principals:

- **Base de dades relacional** que contingui la informació dels buidatges dels fitxers Excel. S'haurà de dissenyar amb els criteris adients per tal emmagatzemar totes les dades disponibles, afavorir consultes SQL que aportin utilitat a les recerques genealògiques i que satisfacin la curiositat dels usuaris en general. Es tractarà d'una base de dades des de la qual es pugui accedir des d'una aplicació web.
- **Aplicació web**, creada amb les tecnologies web actuals, que serveixi de Front-End a la base de dades i permeti la realització de consultes sobre aquesta als seus usuaris. La pàgina disposarà de consultes ja elaborades i parametrizables que facilitin i accelerin les cerques i a més, també proporcionarà estadístiques i anàlisis que puguin ser d'interès.

En relació al contingut, només es tractaran les dades del buidatge proporcionat per l'arxiu de la parròquia d'Agullana, limitant el seu abast. Tampoc es realitzaran més processos de buidatge del que ja s'ha fet, sigui a mà o per mitjà de tècniques de visió per computador, per afegir més registres i ampliar la mostra de dades.

2.1 Obstacles.

La realització d'aquest treball de fi de grau, al igual que en molts altres projectes en general, suposarà l'enfrontament a certs obstacles que s'hauran de superar per garantir l'acompliment dels objectius marcats.

- **Inexperiència i falta de coneixement.** Durant aquest tema es tractaran temes que s'han vist durant la realització del grau, tot i això gran part del seu contingut és nou per l'autor, com per exemple el desenvolupament d'una pàgina web i la connexió amb la base de dades, a part de fer el disseny per una font d'informació bastant gran. Són reptes que s'hauran de superar mitjançant la recerca d'informació i la dedicació constant.
- **Natura de la font de dades.** Les dades que es disposen provenen d'un arxiu que ha passat per les mans de moltes persones, s'ha anat modificant manualment i el seu buidatge també ha estat realitzat a mà per altres persones. El factor humà en el procés de creació de les dades és molt gran, el que pot ocasionar gran quantitat d'errors o incoherències que s'hauran de detectar i neutralitzar per tal de minimitzar el seu impacte en els resultats finals, encara que alguns siguin inevitables.
- **Problemes de privacitat.** Relacionat amb l'obstacle mencionat anteriorment, l'objectiu d'aquest projecte és posar a l'abast de tothom que vulgui una gran base de dades amb avantpassats provinents d'una població catalana, que sense cap dubte està relacionat amb moltes famílies actuals. Això pot derivar en conflictes de privacitat de determinades famílies que no vulguin que certs dades específiques es puguin consultar per tothom, encara que sigui d'avantpassats llunyans.
- **Manca de temps.** Com pot passar en molts projectes, l'acumulació de feina d'altres bandes, els errors de planificació i en les estimacions, així com la mala interpretació d'informació i els problemes de la vida en general, pot provocar que es necessiti treballar més del que es pensava, provocant que s'arribi molt just a etapes finals del projecte. Tot i això aquest obstacle, s'intentarà superar amb èxit gracies al mòdul de gestió de projectes.

2.2 Metodologia de treball.

Per tal de superar els obstacles, minimitzar els efectes dels futurs problemes i tenir un pla a seguir, el projecte s'ha de desenvolupar seguint una metodologia efectiva que doni flexibilitat davant imprevistos però que marqui una guia per evitar perdre's i donar masses voltes. [4]

Per aquesta raó, es proposa una metodologia seqüencial amb fases iteratives:

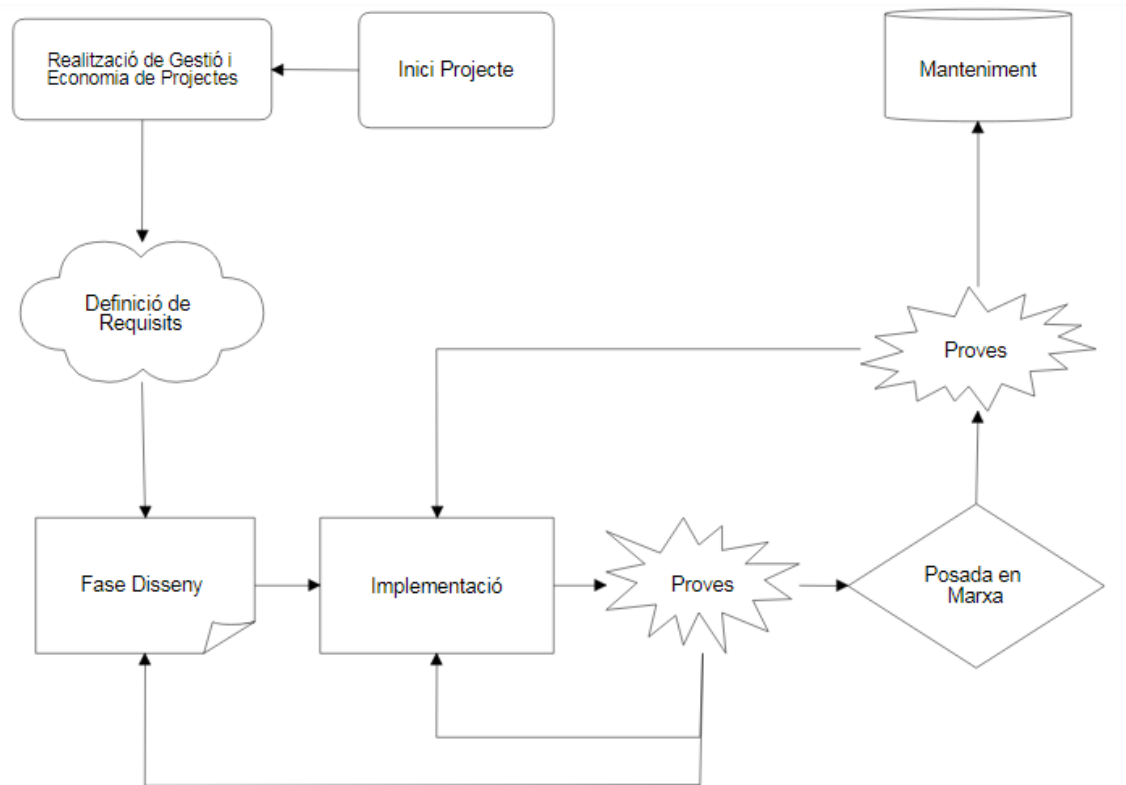


Figura 6. Metodologia triada

Un cop finalitzada l'etapa de GEP, es definirien els requisits a complir del sistema, tant de la base de dades com de l'aplicació web. Seguidament s'entraria en la primera etapa de desenvolupament, que és la del disseny de la base de dades, s'implementaria i es posaria a prova i en funció dels resultats es tornaria a la fase d'implementació o la de disseny. Un cop es posi en funcionament la base dades i es doni el vistiplau, es tornaria a la fase de disseny, però aquest cop per l'aplicació web que seguiria la mateixa metodologia fins l'etapa de manteniment on finalitza el projecte.

2.3 Eines de seguiment i anàlisi.

Per tal de portar un seguiment de les hores total dedicades al treball i analitzar si s'està avançant o no en funció dels temps dedicat i a la planificació, s'utilitzarà l'eina "Toggl", una aplicació web senzilla i fàcil d'utilitzar però efectiva. Aquesta eina permet la creació de projectes amb tasques assignables que tenen una dedicació en hores; simplement quan es comenci a realitzar una tasca s'ha de engegar el cronòmetre i es va registrant el temps. A més, també té funcionalitats de gràfiques on es pot veure l'evolució del temps dedicat a cada tasca i projecte.

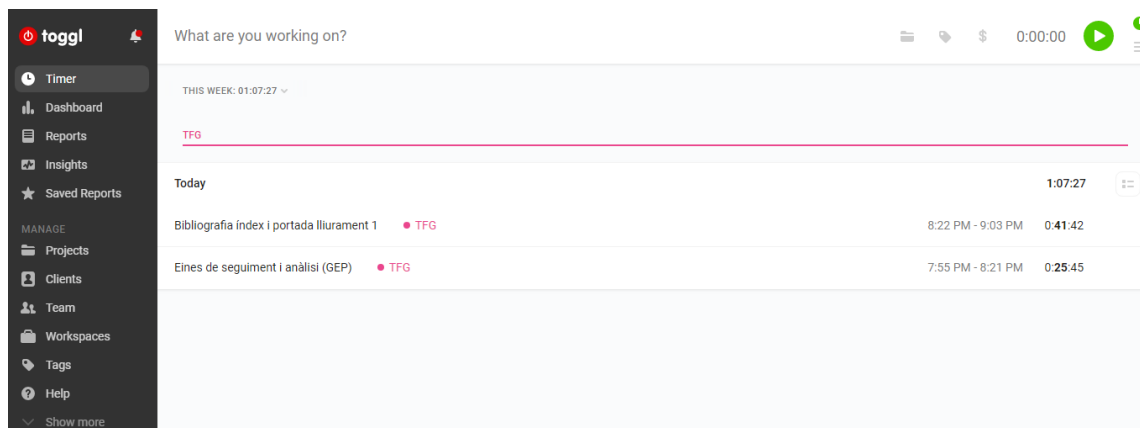


Figura 7. Captura de pantalla de Toggl

Les comunicacions amb el director de projecte es realitzaran majoritàriament per correu electrònic ja sigui per preguntar dubtes, enviar documentació o per programar reunions per orientar el projecte, entre altres aspectes rellevants.

S'utilitzarà Google Drive per emmagatzemar les diferents versions dels lliuraments de GEP i de la memòria final, on cada cop que s'afegeixi i es modifiqui alguna part dels documents es pugi a aquesta aplicació amb l'intenció de disposar de la versió més actualitzada dels documents des de qualsevol dispositiu amb Internet, ja sigui per mòbil, ordinador personal o els ordinadors de la FIB. Aquesta eina també permet anar fent un seguiment del que s'ha escrit cada dia gràcies a que identifica els canvis fets en cada versió. Addicionalment, també s'utilitzarà aquesta eina per l'intercanvi de documents grans (que no es puguin enviar per correu electrònic) entre el director del projecte i l'autor.

Per el desenvolupament de l'aplicació web, s'utilitzarà l'eina GitHub, on anar emmagatzemant les diverses versions de codi que es vagin programant.

3. Planificació temporal

Aquest Treball Final de Grau serà realitzat el quadrimestre de primavera del curs 2017/2018 que transcorre des de la tercera setmana de febrer fins l'última de juny. La última setmana de juny, del 25 al 29, es farà l'última fase, la defensa del projecte, per tant és necessària la realització d'una planificació temporal de les tasques a realitzar per tal de complir tots els objectius abans d'arribar a l'última setmana.

3.1 Descripció de tasques.

Per tal de planificar el transcurs del projecte al llarg del temps disponible per realitzar-lo, s'ha dividit primer en 4 etapes, cadascuna amb les seves tasques.

1ª Etapa

Aquesta part és la destinada al mòdul de Gestió i Economia de Projectes. Està dividida en diverses tasques que corresponen a les diferents entregues que s'han de realitzar. Dura aproximadament un mes.

2ª Etapa

Seguidament, a la segona etapa, cal marcar i especificar clarament quins seran els requisits i casos d'ús del projecte; per continuar amb el disseny de la base de dades i la seva implementació. En aquesta etapa també entra el procés ETL (Extraction, Transport & Load) de migració de les dades, la seva definició i implementació per tal de traslladar el màxim d'informació possible.

3ª Etapa

Un cop la base de dades està creada i omplerta, es pot començar amb la tercera etapa que inclou el disseny i implementació de la pàgina web que funcionarà de Front-End. És una etapa realment important, perquè si no s'adequa als requisits, per molt bé que s'hagi fet l'etapa anterior, pot determinar la qualitat de projecte de cara a l'usuari final.

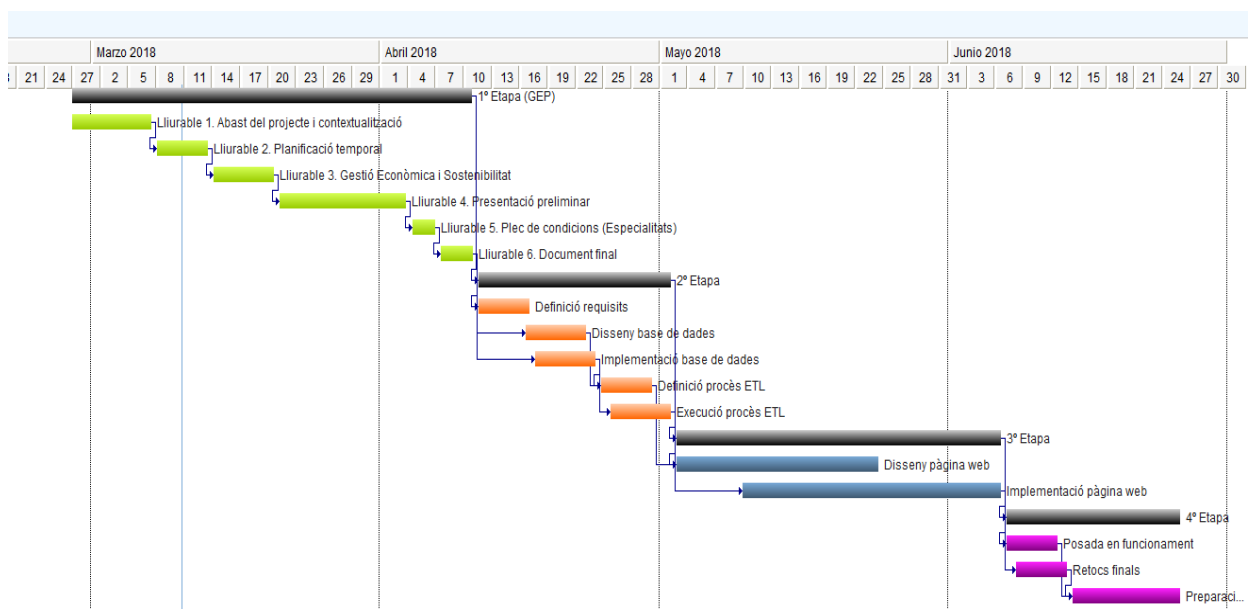
4ª Etapa

L'última etapa està destinada a la posta en funcionament del sistema i als retocs que s'hagin de realitzar per tal d'acabar d'afinar el projecte.

A continuació, s'ha realitzat el diagrama de Gantt d'aquesta planificació per mostrar de forma clara el procediment a seguir i el temps estimat de dedicació per a cada etapa i tasca.

3.2 Diagrama de Gantt

Nombre	Duración	Inicio	Fin	Predecessoras
1º Etapa (GEP)	42días	26/02/2018	09/04/2018	
Lliurable 1. Abast del projecte i contextualització	9días	26/02/2018	06/03/2018	
Lliurable 2. Planificació temporal	5días	07/03/2018	12/03/2018	2
Lliurable 3. Gestió Econòmica i Sostenibilitat	7días	13/03/2018	19/03/2018	3
Lliurable 4. Presentació preliminar	14días	20/03/2018	02/04/2018	4
Lliurable 5. Plec de condicions (Especialitats)	3días	03/04/2018	05/04/2018	5
Lliurable 6. Document final	4días	06/04/2018	09/04/2018	6
2º Etapa	21días	10/04/2018	30/04/2018	1
Definició requisits	6días	10/04/2018	15/04/2018	7
Disseny base de dades	7días	15/04/2018	21/04/2018	7
Implementació base de dades	7días	16/04/2018	22/04/2018	7
Definició procès ETL	6días	23/04/2018	28/04/2018	10,11
Execució procès ETL	7días	24/04/2018	30/04/2018	11
3º Etapa	35días	01/05/2018	04/06/2018	8
Disseny pàgina web	22días	01/05/2018	22/05/2018	12,13
Implementació pàgina web	28días	08/05/2018	04/06/2018	13
4º Etapa	19días	05/06/2018	23/06/2018	14
Posada en funcionament	6días	05/06/2018	10/06/2018	16
Retocs finals	6días	06/06/2018	11/06/2018	16
Preparació defensa	12días	12/06/2018	23/06/2018	18,19



3.3 Distribució del temps i Recursos

Seguint la distribució de temps del diagrama de Gantt anterior i considerant que es treballarà 2,5 hores al dia a l'etapa de GEP i 4 hores al dia per la resta, s'ha elaborat la següent taula on es veu el tant per cent dedicat a cada etapa:

	Temps dedicació(dies)	Hores totals	%
1ª Etapa (GEP)	42	105	25,93%
2ª Etapa (BD)	21	84	20,74%
3ª Etapa (WEB)	35	140	34,57%
4ª Etapa (FINAL)	19	76	18,77%
Total hores	117	405	100,00%

Figura 8. Taula de distribució d'hores

Com es pot veure, l'etapa de l'implementació i disseny de la web és la que portarà més temps, degut al baix coneixement en aquesta tecnologia de l'autor i de la importància que té al projecte.

Ara bé per complir aquesta planificació i distribució de temps es necessiten una sèrie de recursos que portin a terme l'execució del projecte.

3.3.1 Recursos humans

Aquest projecte serà realitzat per un equip desenvolupador de diversos rols que participaran al menys en alguna de les etapes esmentades anteriorment.

- *Cap del projecte*: encarregat de gestionar i planificar el temps, l'economia i la qualitat del projecte.
- *Enginyer de Requisits*: l'encarregat de definir i especificar els requeriments del projecte i les seves parts.
- *Arquitecte Base de Dades*: dissenyador de la base de dades i del procés ETL de migració de dades.
- *Desenvolupador WEB*: implementador i dissenyador de l'aplicació web.

Tots aquests rols seran exercits per l'autor del projecte.

3.3.2 Recursos materials

També seran necessaris determinats recursos materials amb els que es portaran a terme el projecte o donaran suport a la seva execució:

- *Portàtil HP*: Eina per desenvolupar el projecte.
- *PostgreSQL*: sistema de base de dades relacionals molt potent i de codi lliure amb interfícies de programació natives per molts llenguatges de programació.
- *Spyder*: entorn de desenvolupament amb *Python* amb un editor potent i on es desenvoluparà l'aplicació web.
- *Microsoft Office*: software que s'utilitzarà per documentar tot el projecte i la realització de la memòria.

- *Google Drive*: software online on s'emmagatzemaran diversos fitxers i on es guardaran les còpies de seguretat dels lliuraments i de les versions de la memòria.
- *Toggl*: eina per gestionar el temps de les tasques dels projectes.
- *GitHub*: repositori on s'emmagatzemaran les còpies de seguretat del codi desenvolupat.

3.4 Planificació d'alternatives i pla d'acció

A la planificació programada anteriorment, s'ha afegit a l'etapa final la tasca de retocs i també manté el mateix nivell de treball que en les dos etapes anteriors (4 hores al dia), perquè ja es té en consideració la dificultat d'aconseguir que el resultat final, un cop arribada aquesta etapa, sigui l'adequat.

A part d'això, altres problemes poden sorgir, per això és necessari identificar-los abans que apareguin i tenir un pla de contingència per tal de minimitzar els seus efectes.

Possibles Problemes	Probabilitat	Gravetat	Pla de contingència
Falta d'experiència i coneixement	Molt alta	Mitjana	Cercar informació i realitzar cursos online per fer front a la falta de coneixement
Planificació incorrecta	Baixa	Mitjana	Tornar a planificar i dedicar més temps a les tasques en cas de no complir els terminis
Interpretació errònia de la informació	Mitjana	Mitjana	Corregir els errors i dedicant més temps
Averies dels recursos materials	Molt baixa	Alta	Realització de còpies de seguretat dels documents
Accidents laborals o malalties	Molt baixa	Alta	Tenir precaució i hàbits de vida saludables

Figura 9. Taula de possibles problemes i el seu pla de contingència

En general, la majoria de problemes que puguin sorgir es solucionen dedicant més hores al dia i seguint les recomanacions d'experts.

4. Pressupost i Sostenibilitat

4.1. Autoavaluació del domini actual de la competència de sostenibilitat

Un cop realitzat el qüestionari dissenyat per membres del projecte EDINSOST, amb l'objectiu d'analitzar el nivell de formació en sostenibilitat de professors i estudiants del sistema universitari espanyol, he arribat a les següents conclusions:

Al ser un estudiant d'Enginyeria Informàtica d'últim any, sóc ara molt més conscient de l'important paper que juguen les TIC en l'àmbit mediambiental, sociològic i econòmic. Dels avantatges que suposen, com per exemple l'estalvi de tones i tones de paper al emmagatzemar documents en disc durs i el fet de comunicar-se online sense haver-se de desplaçar. Tot i això, també sóc conscient que el fet de fabricar més per vendre més, genera molt rebuig, i en relació a l'àmbit TIC, significa una gran quantitat de residus tecnològics que tenen un impacte molt greu al medi i que s'hauria de fer tot el possible per reutilitzar aquests components mitjançant tècniques de reciclatge. També sé que cal mesurar el consum energètic d'aquests components per tal de reduir-lo i rebaixar al mínim l'impacte a l'ecosistema, al igual que si no es mesura de manera correcta, una solució TIC de vegades pot ser més contaminant que el problema que es pretenia solucionar.

Malgrat el coneixement de les causes i conseqüències de l'impacte dels dispositius electrònics i dels projectes TIC en general a la societat, trobo que no sabia com mesurar-ho d'una manera adequada i realista, degut a la meva creença de que és molt difícil de fer-ho, tot i conèixer mesuradors com per exemple la petjada de carboni, entre d'altres que he vist en assignatures del grau.

En relació als projectes realitzats en l'àmbit TIC, m'agradaria que més d'ells haguessin tingut una finalitat humanitària o mediambiental amb impacte sobre la societat. Dit això, crec que sabia identificar i proposar solucions sostenibles i redirigir projectes cap a un sentit més ecològic i social.

Gràcies a les activitats i tasques realitzades en determinades assignatures del grau que tenien aquest enfocament en sostenibilitat i gestió, em trobo més capacitat per gestionar projectes de manera més sostenible socialment, econòmicament i ecològicament.

4.2. Dimensió Econòmica

En aquest apartat, s'analitzaran i identificaran els diferents costos per estimar el pressupost global d'aquest projecte en qüestió. Es tractaran d'identificar tant els costos directes, indirectes, les amortitzacions, contingències i imprevistos.

4.2.1 Identificació i estimació dels costos

Recursos humans

El projecte, tal i com es va veure a l'apartat anterior de planificació al diagrama de Gantt, consta de diverses etapes en les quals intervindran diferents tipus de perfils, amb un preu per hora determinat per a cadascun, basat en el sou anual de l'Estudi de Remuneració 2017 en Tecnologia de Michael Page [8]. Donat això es calcula el seu import total aproximat:

	Recurs humà	Preu/hora	Total hores	Cost Etapa
Etapa 1 (GEP)	Cap de projecte	21,80 €	105	2.289,00 €
Etapa 2 (BD i Procés ETL)	Enginyer de requisits/Analista	17,50 €	24	420,00 €
	Arquitecte Base de dades	18,40 €	60	1.104,00 €
Etapa 3 (WEB)	Desenvolupador WEB	10,00 €	140	1.400,00 €
Etapa 4 (Retocs i finalització projecte)	Enginyer de qualitat(tester)	12,60 €	38	478,80 €
	Cap de projecte	21,80 €	38	828,40 €
Cost total				6.520,20 €

Figura 10. Cost dels recursos humans

Recursos materials

Per a l'elaboració del projecte, es necessitaran certs recursos materials, tant de software com de hardware, esmentats en anteriors apartats que intervindran en totes les etapes del projecte i que tenen un cert cost i amortització, els quals es detallen a la següent taula:

Recurs	Preu	Hores estimades	Amortització	Cost estimat
Portàtil HP	650 €	405	0,0890 €/hora	36,06 €
PostgreSQL	0 €	60	-	0,00 €
Spyder	0 €	178	-	0,00 €
Windows 7	50 €	405	0,0068 €/hora	2,77 €
Microsoft Office	45 €	143	0,0062 €/hora	0,88 €
Google Drive	0 €	-	-	0,00 €
Toggl	0 €	-	-	0,00 €
GitHub	0 €	-	-	0,00 €
Cost total				39,72 €

Figura 11. Cost dels recursos materials

Per al cost de l'amortització, s'ha considerat que el temps de vida útil dels recursos és de 5 anys i el seu ús de 4 hores diàries, utilitzant la fórmula:

$$\text{Amortització} = \text{Preu} / (5 * 365 * 4)$$

Despeses generals

Ara bé, a l'igual que molts altres projectes, també existeixen certes despeses que afecten indirectament al projecte i que s'han de tenir en consideració per al càlcul del pressupost final. A la següent taula s'especifiquen quins són i el seu cost:

Servei	Preu	Temps	Cost total
Consum energètic	0,15€/kWh	405 hores	15,55 €
Internet	35€/mes	4,5 mesos	157,50 €
Transport	5€/mes	4,5/mesos	22,50 €
Total			195,55 €

Imprevistos

Durant el projecte, poden succeir imprevistos que puguin influenciar el pressupost del projecte, un d'ells és l'averia de l'ordinador portàtil en el qual s'està treballant. Per tal de contrarestar aquest imprevist, s'hauria d'avaluar la gravetat d'aquesta i prendre la millor decisió econòmica i temporal. Considerarem que el preu per reparació es prop del 20% del preu de l'ordinador, és a dir 130€ i amb un risc d'un 5% de probabilitat, amb el que es queda un 6,5€ a considerar al pressupost final.

Pressupost total

Un cop identificades totes les despeses, a la següent taula es pot veure el pressupost total estimat per l'elaboració d'aquest projecte considerant una contingència del 10%:

Tipus	Cost
Recursos humans	6.520,20 €
Recursos materials	39,72 €
Despeses Generals	195,55 €
Cost parcial	6.755,47 €
Contingències	675,54 €
Imprevistos	6,50 €
Total	7.437,51 €

Figura 12. Cost total del projecte

4.2.2 Control de Gestió

Com en molts aspectes de la vida, una cosa és la que es planifica i una altra cosa la que succeeix. Per intentar minimitzar les diferències, o almenys tenir-les controlades al pressupost final del projecte, és necessari el desenvolupament d'un control de gestió.

Per portar-lo a terme, es registraran les hores finalment consumides al projecte al final de cada setmana i al final de cada etapa, d'aquesta manera es podrà fer la comparació entre les hores dedicades i les planificades i d'aquí calcular la desviació entre elles.

Al disposar de l'eina *Toggl*, que va anotant temps dedicat al treball i disposa d'un historial, és molt fàcil registrar les hores finals dedicades realment per a cada tasca.

Llavors es pot fer una taula comparativa que ens permeti calcular les desviacions:

Etap a	Perfil	Hores dedicades	Hores planificades	Desviament hores	Preu final	Preu planifica t	Desviamen t preu
1	Cap de projecte	70	105	35	1526 €	2289€	763€

Figura 13. Exemple taula comparativa

A la taula de dalt, es mostra com es faria el control de gestió un cop acabada la primera etapa, on es pot veure que el temps planificat era més del que finalment ha estat, provocant un estalvi al pressupost final de 763€, el desviament del preu.

4.2.3 Reflexió

El preu estimat final del projecte resulta acceptable per un projecte de software de les seves característiques. S'ha aproximat el cost de cada tasca en funció dels sous dels perfils professionals que hi participen i s'han considerat les despeses en hardware i software així com les més generals. Tot i ser un projecte sense ànim de lucre, perquè el producte final estarà a la disposició de tothom que disposi de connexió a internet, és viable perquè la majoria del cost prové de la part de recursos humans, que seran tots portats a terme per l'autor, el que redueix el cost a les despeses de hardware, software i generals que també seran assumits per l'autor. De cara a la finalització del projecte, si es vol que es posi en funcionament s'haurà de pagar el domini utilitzat així com el servidor, però no es una quota molt alta per tant pot ser assumida per l'autor, per la Facultat, per la Societat Catalana de Genealogia o bé es poden incorporar anuncis a la pàgina de manera que es pugui mantenir per ella mateixa.

El benefici econòmic que tindria la posada en funcionament del projecte seria per la Societat Catalana de Genealogia i els genealogistes catalans, perquè disposarien d'una eina nova de forma gratuïta per als seus estudis de la zona d'Agullana i podria ser un exemple a seguir per a altres projectes que es vulguin realitzar.

En relació al temps i recursos utilitzats en aquest projecte, possiblement en algunes tasques s'utilitzi menys del planificat, però en altres molt més, per tant s'espera arribar a un equilibri que permeti arribar a la fase final de manera còmode i amb gairebé tot enllestit. Potser es podria fer en menys temps si es disposés de persones més experimentades que l'autor en els temes que es tractaran però, amb els recursos humans disponibles, la durada i el material utilitzats són els adequats per assolir els objectius del projecte.

4.3 Dimensió Ambiental

Aquest projecte es realitzarà utilitzant els mínims recursos possibles, ja que per la seva elaboració només es necessari un ordinador, amb connexió a internet quan es necessiti cercar informació i per pujar la pàgina web a la xarxa. L'ordinador que s'utilitzarà es de l'any 2011 i és capaç de suportar la feina que l'exigirà aquest projecte, per tant no es necessari l'adquisició d'un més nou. El transport cap a la universitat quan s'hagin de fer reunions o presentacions presencials es farà caminant, en transport públic o en bicicleta, depenen de les condicions atmosfèriques.

En relació a la energia utilitzada, ja s'ha aproximat en apartats anteriors el seu cost i no baixaria molt el consum en cas d'utilitzar un ordinador més econòmic energèticament.

En relació al que pretén solucionar aquest projecte, permetria estalviar temps d'ús d'ordinadors en les cerques genealògiques perquè aconseguiria fer més ràpid les consultes que anteriorment es feien als fitxers Excel. A part d'això posa a l'abast de tothom aquesta informació per a qui vulgui fer les consultes, estalviant temps de cerca d'aquests fitxers.

4.4 Dimensió Social

Actualment, poca gent coneix ben bé d'on prové ni qui va ser els seus avantpassats llunyans perquè no sap on buscar-ho i qui ho sap, es troba que gairebé molts registres de baptismes i matrimonis, entre d'altres, no estan disponibles per tothom i la major part d'ells en llibres molt antics que es poden fer malbé molt fàcilment. Aquest problema ja l'intenten solucionar organitzacions d'altres països amb major o menor èxit.

Tal i com es va parlar a l'apartat de context i abast del projecte, la genealogia no és una ciència molt coneguda a Catalunya i a la península Ibèrica en general, per això mateix no existeixen les plataformes adequades per realitzar consultes en aquest àmbit que no ocupin molt de temps i dedicació. Per aquest motiu es troba la necessitat d'una eina, on es pugui fer recerca a registres per noms i cognoms entre altres funcions, que ompli aquest buit que hi ha a l'actualitat.

Personalment, aquest projecte m'aporta coneixement sobre un àmbit en el qual m'havia fixat poc, apart dels coneixements tècnics que adquiriré i que em serviran per exercir la professió d'informàtica, i de posar a prova els coneixements que s'han anat aprenent al llarg de tot el grau durant gairebé quatre anys.

Finalment, em quedo amb la visió de l'organització FamilySearch, esmentada anteriorment, que defineix molt bé el perquè de projectes com aquest: "El fet d'aprendre dels nostres avantpassats ens ajuda a entendre millor qui som i a crear un enllaç familiar que uneix el present amb el passat, establint un pont cap al futur"

5. Definició de requisits

Per tal de portar el projecte a bon terme i no desviar-se dels objectius establerts, cal marcar clarament determinats punts que el projecte ha d'assolir per arribar al seu propòsit. El sistema que volem implementar té els seus requisits i es classifiquen de la següent manera [10]:

- **Requisits funcionals:** què ha de fer el software.
- **Requisits de qualitat:** com ha de funcionar el software.
- **Restriccions:** què no ha de fer el software.

A continuació es procedeix a esmentar els requisits trobats:

5.1. Requisits Funcionals

Per a la base de dades:

- Emmagatzemar prop del 100% de les dades dels registres de baptismes, matrimonis i obituaris provinents dels llibres migrats a arxius Excel.
- Identificar les entitats de les dades per crear les taules adequades.
- Mantenir les relacions entre aquestes entitats sense perdre informació.
- Donar accés a les dades per els seus usuaris en qualsevol moment.
- Permetre la connexió amb altres tipus d'aplicacions.
- Donar suport a la creació de còpies de seguretat.

Per a l'aplicació Web:

- Donar accés a les dades dels registres als usuaris habituats a fer servir base de dades.
- Permetre la realització de cerques per nom i cognoms de les persones participants en cada registre per any.
- Permetre la realització de cerques que aportin informació demogràfica.
- Permetre cerques amb finalitats concretes per als estudis genealògics.

5.2. Requisits de qualitat

Per la base de dades:

- Mantenir la integritat de les dades.
- Donar resposta a les consultes en menys de 10 segons.

Per a l'aplicació Web:

- Ha de estar a prova d'errors.
- Ha de seguir un disseny estètic i minimalista per evitar desorientar a l'usuari.
- Els seus components han de ser visibles, consistents i estàndards.

- Ha de permetre tornar a enrere sense ocasionar cap canvi greu.

5.3. Restriccions

- Mantenir la integritat de les dades dels fitxers Excel en la seva migració a la base de dades fent únicament els canvis necessaris per garantir el màxim traspàs d'informació.
- No passar-se de la data límit per la finalització del projecte

6. Disseny de la base de dades

El primer pas que es seguirà per realitzar el disseny de la base de dades és la realització del disseny conceptual en UML (Unified Modelling Language) on es mostraran les classes d'objectes identificades, així com les associacions entre elles i els seus atributs, més les restriccions d'integritat.

6.1. Disseny conceptual en UML

Per identificar les classes d'objectes del nostre sistema són de gran ajuda les capçaleres de les taules dels registres als fitxers Excel, com es pot veure a les següents figures:

A	B	C	D	E	F	G	H
REGISTRE	PERSONA BUIDATGE	PERSONA REVISIO	PARROQUIA	LLIBRE	PAGINA LLIBRE	PAGINA PDF	DATA INSCRIPCIO
1	JOM	PG	Nº Sº de Agullana	B3(1745-1776)2	71v	1	1758/03/03
2	JOM	PG	Nº Sº de Agullana	B3(1745-1776)2	71v	1	1758/03/05

Figura 14. Captura de les capçaleres d'un fitxer de registres

I	J	K	L	M	N	O
NOM OFICIANT	COGNOMS OFICIANT	OCUPACIO OFICIANT	OBS OFICIANT	NOM 1	NOMS COMPLEMENTARIS	COGNOM 1
Julia	Roca	vicari de Agullana		Miquel	Anton, Francisco	Vilar
Francisco	Parer	rector de Agullana		Theodora	Rosa, Maria	Carbonell

Figura 15 Captura de les capçaleres d'un fitxer de registres

En conclusió, cada camp dels fitxers Excel formarà part d'una classe, ja sigui com a classe en si o com a atribut d'aquesta.

Pels registres de baptismes disposem de les següents dades:

- *Dades identificadores del registre*: el número de registre, la persona que va fer el buidatge, la persona que el va revisar, el nom de la parròquia al que pertany (serà sempre la mateixa), el codi identificador del llibre, la pàgina del llibre on està apuntat el registre, la pàgina del PDF del buidatge i la data d'inscripció del registre.
- *Dades identificadores de l'oficiant*: el nom i cognom, així com la seva ocupació i un camp per anotar observacions.

- *Dades identificadores del batejat o batejada:* el nom, els noms complementaris, els dos cognoms, el sexe, la data de naixement i de baptisme, el lloc de naixement i de baptisme i un camp per anotar observacions.
- *Dades de diferents participants a l'esdeveniment:* també tenim informació sobre el pare i la mare dels batejats, així com del avis paterns i materns, dels padrins i del cònjuge de la padrina. La informació d'aquests participants consisteix en majoria del nom i cognoms, lloc de naixement, estat vital, residència i observacions, en cas de ser participants masculins també es disposa de l'ocupació i en cas dels padrins es determina el parentesc.

Per començar amb el disseny, es decideixen les primeres classes, com es pot veure a la següent figura:

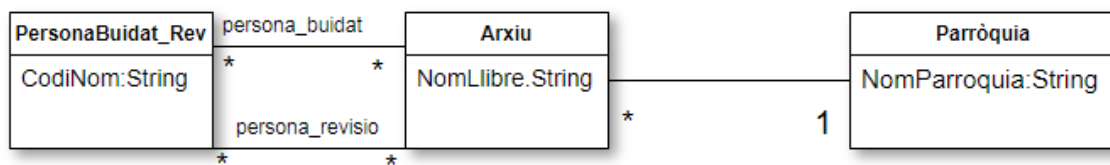


Figura 16. Fragment disseny UML

La classe “Arxiu” està composta pels noms dels llibres, només ha de tenir una sola parròquia assignada i te dos relacions N a N amb la classe “PersonaBuidat_Rev”, on s’emmagatzemarien el codi dels noms de les persones encarregades de buidar i revisar el buidatge del seu document, el que indica que un llibre pot haver estat buidat i revisat per diverses persones.

Per continuar amb el disseny, surten diverses opcions.

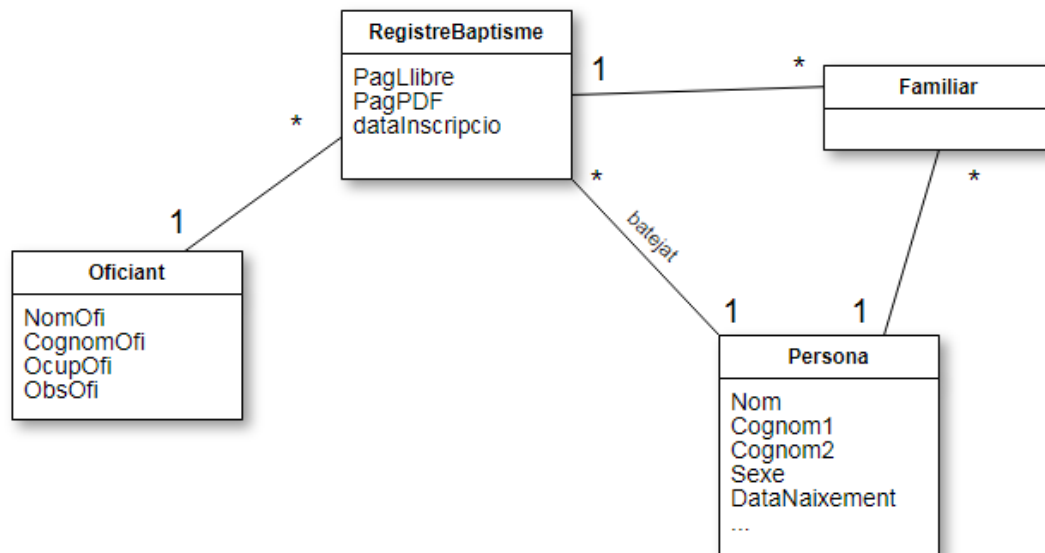


Figura 17. Fragment disseny UML

Una primera opció que es va plantejar en principi pels registres de baptismes és la de l’imatge anterior, on s’identifiquen quatre classes. La primera és la de “RegistreBaptisme”, on es guarden les dades específiques de cada baptisme, la classe

"Oficiant", un dels participants que disposa amb un camp de dades més complert, la classe "Persona", on s'emmagatzemarien totes les dades de les persones participants en un baptisme i finalment la classe "Familiar", que serviria d'enllaç entre persona i registre de baptisme per inscriure la participació d'una persona en un baptisme amb un rol determinat. En relació a aquest rol, es va pensar que es podria determinar quin seria ("Pare", "Mare", "Avi Patern", "Padrí",...) mitjançant subclasses de la classe "Familiar" que incorporessin les dades específiques de cada rol (per exemple Padrí conté el camp parentesc, que no està en altres participants perquè el parentesc en si indica el rol que fan).

La classe "Oficiant" es podria separar de "Familiar" perquè el seu rol és més important en relació al registre de baptisme i al contrari que els altres, les seves dades estan molt més complertes, possiblement perquè l'oficiant era o bé l'encarregat d'escriure manualment el registre o bé el superior d'aquell qui ho feia perquè al cap i a la fi era qui batejava al nadó. Però en relació a la genealogia, el seu interès no és tan gran com per fer una classe extra, per tant es podria incloure també a la taula "Familiar", que llavors es convertirà en la classe "Participant".

La decisió de fer subclasses de la classe "Familiar", ara "Participant", per identificar el rol del participant, suposaria un gran increment de taules a la base de dades definitiva, per això finalment es decideix crear un "enumeration" com a atribut de la classe participant que contingui els diferents tipus de participant que pot haver, més el camp d'observacions individual per a cada participant.

Ara bé, com també disposem de registres de matrimonis i d'obituaris, que aportarien molt valor genealògic a futures consultes, s'han de considerar també al disseny conceptual encara que no se'n disposi de tants com de baptismes. Per incloure'ls, es podrien crear dues taules més similars a la de "registreBaptisme", una per matrimonis i l'altra per obituaris, però com tots tres tipus d'esdeveniments comparteixen camps de dades similars es decideix crear una classe "Esdeveniment" que contingui camps en comú com la data d'inscripció i l'observació general. D'aquesta nova classe, es generarien tres subclasses que es correspondrien amb "Baptisme", "Matrimoni" i "Obituari" amb els seus camps de dades particulars cadascuna. A més, serà necessari afegir l'atribut "tipus", un altre "enumeration", a la classe "Esdeveniment", que identifiqui quin tipus d'esdeveniment és i es crearà una altra classe anomenada registre, que contingui els atributs "número registre", "pàgina de llibre" i pàgina PDF" perquè no estan relacionades amb l'esdeveniment específicament.

A part, es decideix crear una "enumeration" per cadascun dels camps que tenen una quantitat limitada de valors, com per exemple l'estat vital (viu/va o difunt/a).

Dit això, quedaria un disseny UML segons l'imatge següent:

En relació a les restriccions d'integritat, no hi haurien moltes, perquè al ser dades provinents de registres escrits a mà per diverses persones durant un període de temps molt gran, hi ha molta probabilitat de que faltin dades. Per exemple, per un baptisme no es pot assegurar que un nadó tingui pare i mare, perquè hi ha casos en que un d'ells falta o fins i tot en ocasions tots dos, o tampoc es pot establir que un registre ha de tenir un i només un oficiant perquè també hi ha registres en que no s'indica aquest camp. Per això és preferible no establir restriccions que puguin ocasionar pèrdues de dades de cara a les consultes finals, a més que la falta d'aquestes restriccions no afectaria gaire al resultat de les consultes. També s'ha de considerar que no s'espera inserir més dades un cop s'hagi fet la migració, per això no seran necessàries les restriccions en aquest cas.

Un altre problema que també s'haurà de tractar per culpa de la natura de les dades és l'identificació de les classes per clau primària. Per exemple, és impossible identificar a una persona fent que les claus siguin el nom i els cognoms, primerament perquè per gran part de les persones el segon cognom no s'indica i hi ha molts casos en que falta el nom o fins tot el primer cognom. Ignorar aquestes dades provocaria una gran pèrdua d'informació, a més que tampoc es garanteix que siguin úniques ja que al ser dades generades durant un període molt gran de temps es pot donar el cas de persones que es diguin exactament igual. Tampoc es podria utilitzar com a clau cap altre camp per la mateixa raó que no és únic i sempre es dona el cas de que falta per un registre determinat. Dit això, la solució podria ser crear una taula amb una clau que es generi automàticament a mesura que es van inserint a la taula o bé simplement no especificar cap clau primària. Finalment, es pren la decisió de crear la taula sense clau primària perquè la clau automàtica tampoc aportaria gaire a les consultes.

Després d'aquests raonaments obtenim les següents restriccions d'integritat:

1. Claus primàries: `PersonaBuidat_Rev(CodiNom)`, `Parròquia(NomParroquia)`, `Arxiu(NomLLibre)`, `Registre(NumRegistre)`.
2. En un mateix arxiu, `persona_buidat` ha de ser diferent a `persona_revisio`.

El següent pas es crear les capçaleres de les taules en funció de les classes i associacions del diagrama UML.

6.2. Disseny de les taules de la BBDD

Abans de traduir un disseny UML a taules d'una base de dades, cal seguir una sèries de passos per la seva traducció.

Primer de tot cal eliminar les associacions en les que intervinguin més de dos classes, les classes associatives (classes generades a partir d'una associació entre classes), controlar les restriccions d'integritat i tractar la informació derivada (atributs calculats). Però com no es produeixen cap d'aquests casos, podem procedir amb la traducció.[12][13]

Es decideix començar la traducció per les dos taules més senzilles que tenim que són la que emmagatzema el codi de les persones que fan el buidatge o revisen i el de la parròquia.

- Persona_Buidat_Rev (**Codi_Nom**)
- Parròquia (**Nom_Parroquia**)

Es marquen en negreta les que serien les claus primàries de les taules. Per traduir la taula Arxiu, s'ha especificat al UML que un arxiu ha d'incloure una i només una parròquia i pot tenir més d'una persona que hagi fet el buidatge, per tant ha d'incloure claus primàries d'altres taules com a columnes de la seva:

- Arxiu (**Nom_Llibre**, **Nom_Parroquia**, P_Buidat, P_Revisat)

{Nom_Parroquia} references Parròquia

{P_Buidat, P_revisat} references Persona_Buidat_Rev

En aquest cas s'indica amb subratllat les claus foranes i s'especifica d'on provenen a sota. Donat això es procedeix per fer les traduccions de la resta de les taules.

- Registre (**Nom_Llibre**, **Num_registre**, Pag_Llibre, PagPDF)

{Nom_Llibre} references Arxiu

- Esdeveniment (**Nom_Llibre**, **Num_Registre**, Data_Inscripcio, DataEsdeveniment, LlocEsdeveniment, tipus, ObservacióGeneral)

{Nom_Llibre, Num_Registre} references Registre

- Persona (Nom, Cognom1, Cognom2, Sexe, Estat_vital, ofici_carrec, alies, Estat_civil, data_naixement, lloc_naixement, residència)

- Baptisme (**Nom_Llibre**, **Num_Registre**, Nom, Noms_complementaris, Cognom1, Cognom2, sexe, data_naixement, lloc_naixement, observacio)

{Nom_Llibre, Num_Registre} references Esdeveniment

{Nom, Cognom1, Cognom2, sexe, data_naixement, lloc_naixement} references Persona

- Matrimoni(**Nom_Llibre**, **Num_Registre**, Nom_Marit, Cognom1_Marit, Cognom2_Marit, Edat_Marit, Lloc_naixement_Marit, Estat_civil_Marit, residència_Marit, Ocupacio_Marit, Nom_Muller, Cognom1_Muller, Cognom2_Muller, Alies_Muller, Lloc_naixement_Muller, residència_Muller, Edat_Muller, Estat_civil_Muller, Església, Capitols_Mat)

{Nom_Llibre, Num_Registre} references Esdeveniment

{Nom_Marit, Cognom1_Marit, Cognom2_Marit, Lloc_naixement_Marit, Estat_civil_Marit, residencia_Marit, Ocupacio_Marit} references Persona
{Nom_Muller, Cognom1_Muller, Cognom2_Muller, Lloc_naixement_Muller, Estat_civil_Muller, residencia_Muller, Alies_Muller} references Persona

- Obituari(**Nom_Llibre, Num_Registre**, Nom, Cognom1, Cognom2, alies, sexe, data_naixement, lloc_naixement, edat, estat_civil, residencia, ocupacio, lloc_defuncio, cementiri, noms_fills)

{Nom_Llibre, Num_Registre} references Esdeveniment

{Nom, Cognom1, Cognom2, alies, sexe, data_naixement, lloc_naixement, estat_civil, residencia, ocupacio} references Persona

- Participant (**Nom_Llibre, Num_Registre**, tipus, tipus_Familia, Nom, Cognom1, Cognom2, Estat_vital, ofici_Carrec, lloc_naixement, residencia, Observacions)

{Nom_Llibre, Num_Registre} references Esdeveniment

Per traduir les subclasses Baptisme, Matrimoni i Obituari, que provenen de la classe "Esdeveniment", es decideix seguir l'estratègia "Class Table Inheritance", en la qual per a cada classe del diagrama es crea una taula. Es podria haver seguit l'estratègia "Concrete Table Inheritance" on les dades de la classe Esdeveniment s'incloïen dins de les subclasses i d'aquesta manera s'estalviava la generació d'aquesta taula, però convé separar-la per facilitar les consultes amb unions amb la classe participant.

6.3. Elecció del Sistema Gestor de Base de Dades

Hi ha molts sistemes gestors de bases de dades per a models relacionals que són de software lliure, com per exemple MariaDB o MySQL, però finalment s'ha triat PostgreSQL perquè compleix de forma excel·lent les exigències d'aquest projecte.

Per PostgreSQL, gràcies a ser de software lliure, és molt més fàcil trobar documentació online d'altres desenvolupadors que ajudin a resoldre problemes que puguin sorgir. A més, l'equip encarregat del seu desenvolupament és actiu i acostuma a realitzar actualitzacions freqüents del software per millorar l'experiència d'usuari i augmentar les capacitats d'aquest gestor. [11]

El procés d'extracció de dades es farà amb *Python*, llenguatge de programació que proporciona bones llibreries per connectar-se amb PostgreSQL i del qual hi ha molta documentació a Internet de com realitzar-la. Si s'hagués triat Oracle per exemple, no hi hauria tant codi lliure per documentar-se.

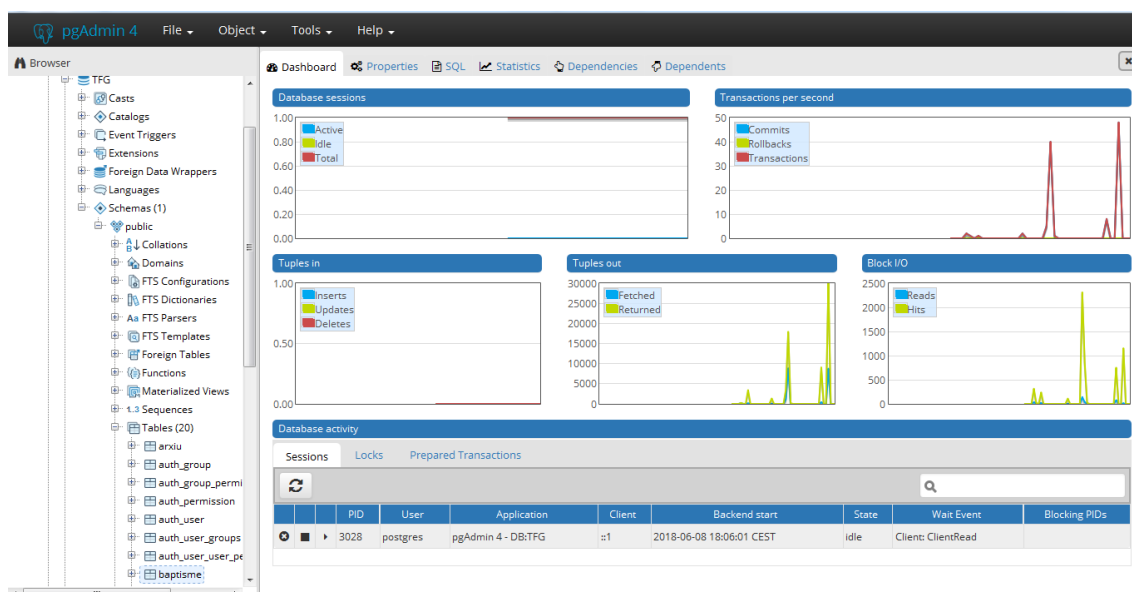


Figura 19. Captura de pantalla de pgAdmin4 per PostgreSQL

Es pot consultar l'script utilitzat per la creació de les taules a PostgreSQL a l'Annex.

7. Procés Extraction-Transform-Load (ETL)

El procés ETL (Extraure, Transformar i Carregar) és el que s'encarrega de moure dades provinents de múltiples fonts, donar-les-hi format i netejar-les per després carregar-les a una base, un magatzem o un repositori de dades. En aquest cas, el que volem és agafar les dades dels fitxers Excel i aplicar els canvis que siguin necessaris al format per adaptar-les als tipus de dades de la base de dades creada al punt anterior. Per portar a terme aquest procés, es crearà un script en *Python*, llenguatge de programació que proporciona les llibreries necessàries per treballar amb fitxers Excel i amb PostgreSQL.

Les llibreries més importants per a la realització d'aquest procés són “pandas”, de codi obert amb llicència BSD (Berkeley Software Distribution), que proporciona estructures de dades d'alt rendiment fàcils d'usar i eines d'anàlisi de dades, i la llibreria “psycopg2”, la més popular per connectar-se a PostgreSQL per fer consultes, insercions i actualitzacions des de *Python*. [14][15]

7.1. Mètode seguit

Amb l'objectiu de ser consistent i d'obtenir el màxim d'informació possible, s'organitza un mètode que es pugui programar en un script de *Python*. El mètode implementat és el següent:

1. S'organitzen els fitxers Excel en tres carpetes, una amb els registres de baptismes, que s'identifiquen perquè són els que comencen amb la lletra “B”, una per els matrimonis, que s'identifiquen perquè són els que comencen amb la lletra “M” i una altra per obituaris, que comencen amb “O”. A l'script de *Python*, es creen les variables que representin cadascuna de les taules de la base de dades, és a dir, per Persona_Buidat_Rev, Parròquia, Arxiu, Registre, Esdeveniment, Persona, Baptisme, Matrimoni, Obituari i Participant.

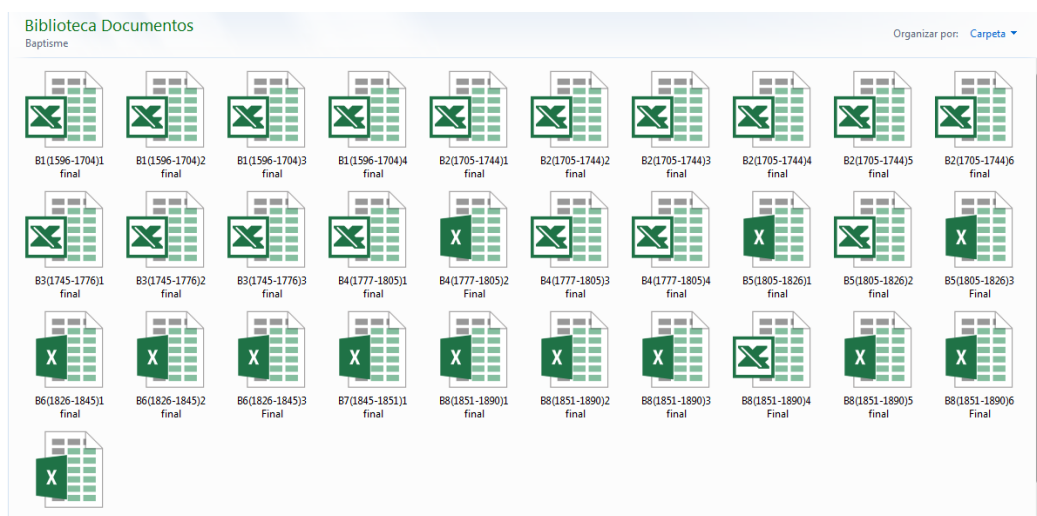


Figura 20. Captura on es veuen tots els fitxers de Baptismes a la seva carpeta

2. Per a cadascun d'aquests directoris, es recorren tots els fitxers i es carreguen utilitzant la funció "ExcelFile" de "pandas", seleccionant pel seu nom el full de càlcul a utilitzar amb "parse" per obtenir totes les dades en un "DataFrame", objecte que utilitza "pandas" per emmagatzemar les matrius de dades. Prèviament, ens hem d'assegurar que tots els fulls a utilitzar dins del fitxers Excel tinguin el mateix nom per automatitzar el procés.

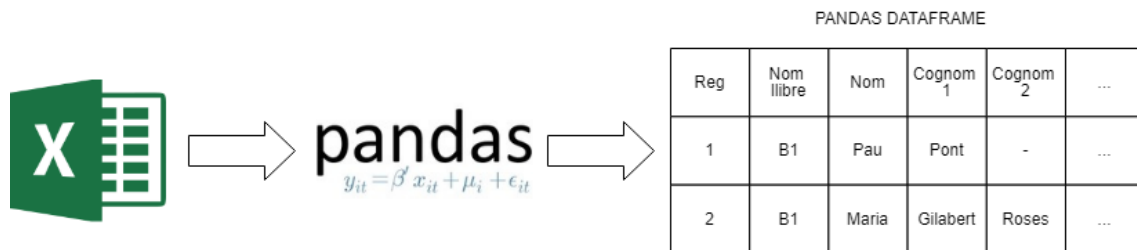


Figura 21. Esquema del procés E.T.L.

3. Un cop tenim les dades al "DataFrame", es procedeix a recorre cada una de les files d'aquesta matriu i a emmagatzemar cada columna en una variable per ser tractada posteriorment. Per exemple, per a un baptisme es crearia una variable pel número de registre, per cadascuna de les dates, per a cada nom i cognom dels participants, etc. Si un fitxer te 70 columnes amb dades, es crearien 70 variables. Això es fa per facilitar el procés de tractament de les dades abans d'inserir-les a la base perquè és més fàcil i es redueixen els errors al tractar amb el nom d'allò que representa el valor, per exemple s'entén més què està referenciant la variable "NumRegistre", que la comanda d'accés a un valor de la matriu ("df.get_value(0,0)" en el cas de consultar el valor de la fila 0 a la columna 0 del "DataFrame" df).
4. Seguidament, s'insereixen a cada taula de les creades a l'script del pas 1 (no a la base de dades) els valors necessaris, és a dir, s'insereixen les dades de la Parròquia a la taula Parròquia, les dades del protagonista a les taules Persona i Baptisme (en el cas dels baptismes), les dades de l'oficiant, pare, mare o qualsevol altre participant a la taula Persona i Participant, etc. La inserció a aquestes taules està sotmesa a certes condicions que simulen les restriccions que tindrien a una base de dades per després facilitar la inserció definitiva a la base sense que salti cap tipus d'error o excepció. Aquestes condicions seran explicades més endavant.

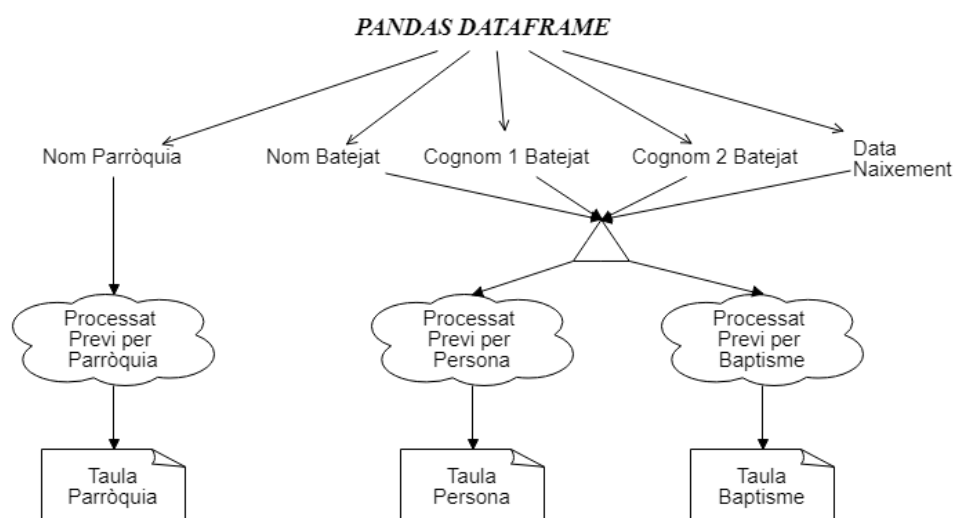


Figura 22. Esquema del procés E.T.L. que resumeix els punts 3 i 4.

5. A continuació, arriba l'últim pas que és la connexió i la inserció de les dades a la base. La funció "connect" de la llibreria "psycopg2" permet fer la connexió passant-li com a paràmetres el nom de la base de dades, l'usuari, el host i la contrasenya. Per poder fer les insercions, l'objecte que ens dona la funció "connect" ens proporciona un cursor des d'on executar les comandes a la base de dades. De fet, ens dona la possibilitat d'inserir una a una les fileres a cada base amb la comanda "execute" però també podem inserir-les totes amb una sola comanda utilitzant la comanda "executemany" si disposem de totes les dades en taules, com és el cas d'aquest mètode.

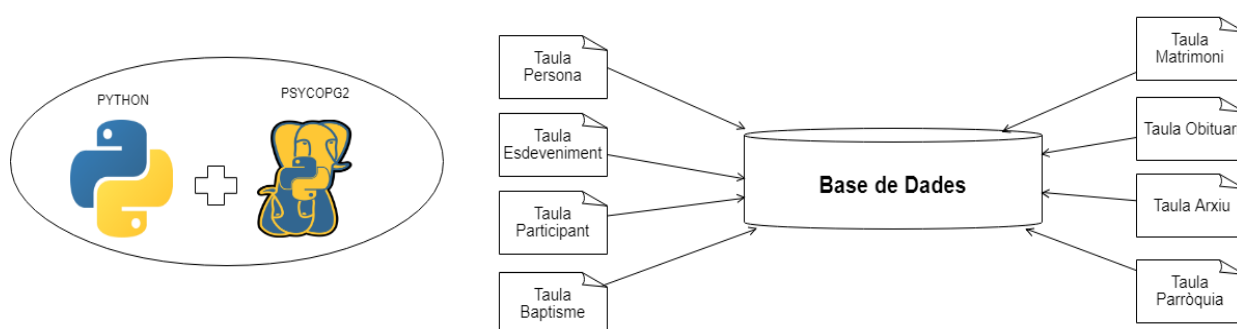


Figura 23. Esquema del procés E.T.L.

Utilitzant aquest mètode s'aconsegueix fer el procés ETL clau d'aquest projecte. Aproximadament, el tractament de les dades te una durada d'execució de 20 minuts i inserir les files a les taules de la base dades amb "psycopg2" triga al voltant de 5 minuts, el que demostra la gran eficiència d'aquest mòdul.

Tot aquest procés es pot veure implementat en *Python* a l'Annex.

7.2. Dificultats trobades i condicions a seguir en el procés ETL

Les dades dels fitxers Excels on es troben els registres tenen un patró que segueixen fidelment per a les columnes de cada tipus de registre (Baptisme, Matrimoni o Obituari), malgrat això hi ha variacions i errors a les columnes que fan que s'hagin de tractar abans de inserir-les a la base de dades. Aquests errors són a causa de la natura de la font de dades perquè provenen d'escrits realitzats a mà en documents antics i migrats, també a mà, a fitxers Excel. Probablement les persones que van fer la migració van aconseguir reduir aquests errors i variacions de format de manera considerable però, tot i així, encara queden molts que han de ser tractats.

Un dels errors que afectaven més el disseny d'aquest sistema de base de dades és el dels números identificadors dels registres, camp clau per inserir i relacionar diverses taules entre elles. Durant la realització del procés ETL, en accedir al camp dels números de registres, de vegades es trobaven números en blanc o que estaven repetits en un mateix llibre o que tenien caràcters que no es corresponien als números enters, ocasionant problemes d'integritat. Per evitar aquest error, es va haver de tornar a calcular aquest camp per a cadascun dels fitxers, garantint que no es violaria cap de les restriccions establertes per aquest camp.

Abans d'inserir res a la base de dades, primer s'han d'afegir els valors corresponents a les variables de l'script que representen cada taula a la base, tal i com s'ha vist a l'apartat anterior. Prèviament a introduir una fila a alguna d'aquestes taules, s'han de comprovar una sèrie de condicions que mantinguin la consistència i format necessaris per tal d'evitar errors i facilitar la posterior consulta en etapes posteriors del projecte.

Totes les dades, siguin del camp que siguin, han de passar per un processat previ:

- Si la dada és de format "unicode": s'ha de codificar al format "utf-8" i eliminar els espais en blanc innecessaris als laterals amb la funció per a strings de *Python* "strip()".
- Si la dada és la cadena de caràcters "NN", "(NN)", "x" o "": canviar el valor de la dada pel format nul de *Python* "None".
- Si la dada és del tipus float o de tipus numpy.float64 i és nul·la: canviar el valor de la dada pel format nul de *Python* "None".
- Si la dada és del tipus numpy.float64: canviar el tipus de dada a l'enter de *Python*.

Aquest processat previ es realitza per garantir el mateix tipus de dades per a cada columna i per afegir el valor nul "None" a aquells camps buits o indicadors de valors no trobats, els quals després serà necessari que estiguin en aquest format per inserir-los a la base de dades.

A continuació, es descriu una sèrie de condicions que no es fan en el processat previ i que s'han de comprovar per a cada taula:

Persona_Buitat_Rev. Persones que han buidat o revisat els fitxers.

- La fila pot ser inserida si el valor de la dada no és nul, si no estava ja inserit a la taula i si te una longitud menor o igual a 4. Es comprova la longitud atès que al ser un codi dels noms de les persones acostuma a ser breu, com per exemple les inicials del nom complert, per això si és més gran de 4 probablement sigui un error.

Parròquia. Nom de la parròquia on es va fer el baptisme.

- La fila pot ser inserida si el valor de la dada no és nul i no estava ja inserit a la taula.

Arxiu. Nom de l'arxiu, la parròquia i les persones que van fer el buidatge.

- La fila pot ser inserida si el nom del llibre no és nul i no estava ja inserit a la taula.

Registre. Número de registre, pàgina del llibre i de l'arxiu pdf on està localitzat el registre.

- La fila pot ser inserida si la parella nom del llibre i número de registre no te algun valor nul i no estava ja inserida a la taula.
- Si el tipus de la pàgina del PDF no és enter, es posa el valor a nul. La pàgina del llibre no es comprova perquè es una cadena de caràcters.

Esdeveniment. Indica les dates, el lloc i el tipus d'esdeveniment més les observacions generals del registre.

- La fila pot ser inserida si la parella nom del llibre i número de registre no te algun valor nul i no estava ja inserida a la taula.
- Si la data d'inscripció o la data de l'esdeveniment no tenen valor nul i tenen una llargària més gran de 12, es posa el valor a nul. Les dates s'ha decidit tractar-les com a cadenes de caràcters i no com el tipus data de *Python*, que s'adaptaria al tipus data del SQL, perquè es donen molts casos en els quals falta el dia o falta el mes o apareixen símbols d'interrogació indicant incertesa sobre alguna part de la data... Això afegirà dificultat a l'hora de fer consultes en les quals les dates tinguin un paper important. En relació a la longitud de les dates, s'ha especificat que es posin a nul sí tenen una llargària més gran de 12 perquè normalment la longitud d'una data és de 10 caràcters (DD/MM/YYYY) i es tolera que es puguin afegir dos símbols extra, com per exemple els interrogants.
- Si el camp d'observació general no és nul i te una longitud més gran de 500 caràcters, només es consideren els primers 500. Es pren aquesta mesura per evitar tenir camps amb molta lletra al resultat de les consultes.

Persona. Emmagatzema diverses dades sobre una persona.

- La fila pot ser inserida si almenys el nom o algun dels cognoms no és nul i si no hi ha cap altra fila que tingui exactament igual el conjunt de nom, cognoms, estat vital, data naixement i lloc de naixement.
- Si la data de naixement no te un valor nul i la seva llargària és més de 12 caràcters, es posa el valor a nul.
- Si el primer cognom conté el símbol "+", es canvia l'estat vital a difunt.

- Si el nom, els cognoms o l'estat civil conté més de 50 caràcters, només es prenen els primers 50 caràcters per al camp corresponent. Es pren aquesta mesura per evitar que apareguin anotacions que haurien d'estar al camp d'observacions i per millorar la relació entre possibles consultes posteriors que incloguin el nom i els cognoms.

Participant. Emmagatzema les dades d'un participant a un esdeveniment indicant quin es el rol que realitza dins d'ell.

- La fila pot ser inserida si la parella nom del llibre i número de registre no te algun valor nul i no estava ja inserida a la taula.
- Si el primer cognom conté el símbol "+", es canvia l'estat vital a difunt.
- Si el nom, els cognoms o l'estat civil conté més de 50 caràcters, només es prenen els primers 50 caràcters per al camp corresponent

Baptismes, Matrimonis i Obituaris. Indica les dades del nadó, de la parella o del difunt, respectivament.

- La fila pot ser inserida si la parella nom del llibre i número de registre no te algun valor nul i no estava ja inserida a la taula.
- Si el nom, els cognoms o l'estat civil conté més de 50 caràcters, només es prenen els primers 50 caràcters per al camp corresponent
- Si el camp d'observacions no és nul i te una longitud més gran de 500 caràcters, només es consideren els primers 500.

Un cop considerades aquestes condicions per a totes les dades i després d'haver estat aquestes dades inserides a la base, ja està complerta una part del projecte i es pot procedir al disseny i implementació de l'aplicació web des d'on s'accediran a les dades i es podran fer les consultes.

8. Disseny i Implementació de l'aplicació web

La segona part d'aquest projecte consisteix en el disseny d'una aplicació web des de la qual fer accessibles les dades i l'informació del registres que hi ha als fitxers Excel.

Actualment, hi ha moltes eines que faciliten la creació de pàgines web. Abans de començar el desenvolupament, s'ha de triar si es faria directament programant des de zero o s'utilitzaria un framework que accelerés el procés. Efectivament, aquest últim punt seria l'adequat a causa del poc coneixement en tecnologies web de l'autor i donaria una empenta al desenvolupament. Com l'etapa anterior del projecte, el procés ETL, s'ha fet utilitzant *Python*, seria un punt a favor trobar un entorn que permetés crear l'aplicació web utilitzant el mateix llenguatge. Amb aquesta característica, s'ha escollit el framework *Django*.

8.1. Framework escollit

L'escollit per la realització de l'aplicació web ha estat *Django*, un framework per aplicacions web d'alt nivell que afavoreix el ràpid desenvolupament perquè s'ocupa d'automatitzar moltes molèsties del disseny web. [16]

Característiques de *Django*:

- Rapidesa. *Django* ha estat dissenyat per ajudar als desenvolupadors a portar les aplicacions del concepte al producte final de la manera més ràpida.
- Ben equipat. Inclou molts extres que ajuden a manejar tasques comunes del desenvolupament web.
- Seguretat. Ajuda als desenvolupadors a evitar errors comuns de seguretat.
- Escalable.
- Versàtil. Es poden crear molts tipus d'aplicacions diferents.

Pel desenvolupament del projecte ens interessa bàsicament que sigui ràpid, versàtil i ben equipat per complir els requisits.[18]



Figura 24. Logo Django. Font: djangoproject.com

8.2. Primers passos de l'aplicació web

Un cop instal·lat el *Django*, es pot començar amb l'implementació de l'arrel de la web abans de començar amb el disseny.

Una de les facilitats de *Django* és que amb l'execució d'una comanda a la consola del sistema crea l'estructura d'arxius necessària per començar el projecte. La comanda és la següent:

```
> django-admin startproject mysite
```

Figura 25. Comanda per crear la base de la web

Aquesta comanda crea una carpeta, anomenada “mysite” en aquest cas però pot tenir una altre nom, que contindrà tots els components del projecte. Dins d'aquesta carpeta es creen:

- El fitxer “manage.py”, un programa automàtic generat per interactuar amb el projecte de diferents maneres, com per exemple per llançar l'aplicació.
- Un altre directori anomenat “mysite” que, al contrari que el directori que el conté, no pot canviar de nom i que a la vegada conté altres components necessaris de projecte:
 - El fitxer “__init__.py” que indica a *Python* que el directori en concret és un “Package”, és a dir, un mòdul per poder utilitzar-lo des d'altres fitxers.
 - El fitxer “settings.py”, on s'especifiquen algunes constants necessàries per la aplicació, com per exemple els paràmetres de connexió amb la base de dades.
 - El fitxer “urls.py”, on s'indiquen totes les URL del projecte, és a dir, les adreces de les pàgines web.

Aquests serien els components bàsics que tindria qualsevol aplicació creada amb *Django*. Ara bé, per desenvolupar l'aplicació cal crear un altre directori dins de la carpeta mysite, on es troba el fitxer “manage.py”. Aquest directori es crea amb una comanda diferent a l'anterior:

```
\mysite> python manage.py startapp tfg
```

Figura 26. Comanda per crear l'aplicació

D'aquesta manera es crea l'aplicació “tfg”, que serà un altre directori de fitxers que contindrà, entre d'altres, els següents fitxers:

- El fitxer “__init__.py” que indica a *Python* que el directori en concret és un “Package”, és a dir, un mòdul per poder utilitzar-lo des d'altres fitxers.
- El fitxer “app.py” que indica al framework que el directori en concret és una aplicació del projecte.
- El fitxer “urls.py”, on s'indica totes les URL de l'aplicació i quina funció s'executa per a cadascuna.
- El fitxer “views.py”, on es programen les accions que ha de realitzar l'aplicació en funció de l'URL demanada.

- El fitxer “models.py”, on s’especifiquen els models de objectes que tractarà l’aplicació.
- Un directori anomenat “migrations”, on es registraran els canvis fets a la base de dades del projecte.

Aquest projecte de *Django* utilitza com a programa principal el contingut que hi hagi a la carpeta “mysite” per això és important enllaçar-la amb l’aplicació “tfg” creada. Per aconseguir-ho, cal afegir al fitxer “urls.py” de “mysite” un paràmetre que indiqui que també utilitzi els URL de “tfg”:

```
urlpatterns = [
    path('tfg/', include('tfg.urls')),
    path('admin/', admin.site.urls),
]
```

Figura 27. Fragment de codi del fitxer url.py de mysite

A més, també cal modificar del fitxer “settings.py” la variable “INSTALLED APPS” i afegir “tfg” a la llista d’aplicacions utilitzades.

Un altre pas clau pel funcionament de l’aplicació, és la connexió amb la base de dades. Per realitzar això, també s’ha de modificar una variable del fitxer “settings.py”, aquest cop la variable DATABASES, on s’ha d’especificar almenys la base de dades per defecte. En aquest cas, s’utilitzarà la creada prèviament on tenim tota l’informació dels registres de baptismes. El framework crearà les taules necessàries pel funcionament de l’aplicació i ja tindrà accés a les taules dels registres. Per realitzar aquesta connexió, cal especificar el nom de la base de dades, l’usuari, la contrasenya, el host, el port i el motor utilitzat per *Django* per fer aquesta connexió, tal i com es pot veure en el següent fragment de codi:

```
DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.postgresql_psycopg2',
        'NAME': 'TFG',
        'USER': 'postgres',
        'PASSWORD': 'peca2012',
        'HOST': 'localhost',
        'PORT': '',
    }
}
```

Figura 28. Fragment de codi de settings.py

Adicionalment, un dels beneficis que aporta *Django* a aquest projecte és que te molt bona compatibilitat amb el sistema gestor de base de dades utilitzat, PostgreSQL, de fet és el recomanat pels desenvolupadors del framework.

Continuant amb més característiques, *Django* ofereix la possibilitat de generar el que s’anomena models, objectes que contenen tots els camps essencials i el comportament

de les dades emmagatzemades. En altres paraules, amb aquesta funcionalitat de *Django* es podria representar les taules creades prèviament a la base de dades i utilitzar-les per fer les consultes sobre elles. En el cas d'utilitzar només aquest sistema, s'estaria seguint el principi DRY (Don't Repeat Yourself) que diu que cada concepte i/o part de dades hauria d'estar present en un i només un lloc, evitant redundàncies. Tot i això, utilitzar aquest sistema no seria adequat pel projecte perquè ja tenim les taules creades i el seu sistema d'accés a les dades no es tant potent com l'SQL.

```
class Registre(models.Model):
    nom_llibre = models.CharField(max_length=20)
    num_registre = models.IntegerField()
    pag_llibre = models.CharField(max_length=20, blank=True, null=True)
    pag_pdf = models.IntegerField(blank=True, null=True)
    objects = CopyManager()

    class Meta:
        managed = True
        db_table = 'registre'
        unique_together = (('nom_llibre', 'num_registre'),)
```

Figura 29. Exemple de com es representaria un Registre segons els models de Django

8.3. Creació i disseny d'una pàgina web

Abans de la reproducció d'una pàgina web en un navegador, es recorren una sèrie de passos previs que s'han de tenir en compte a l'hora d'implementar l'aplicació.

Primer de tot, quan des del navegador es demana una certa URL al servidor, aquest la busca entre un llistat d'URL possibles. Al projecte, aquestes URL es troben al fitxer "urls.py" juntament amb l'acció que han d'executar. Aquesta acció habitualment consisteix en cridar una funció, localitzada al fitxer "views.py", que retorni una resposta HTTP que torni un arxiu HTML amb els càlculs programats.

Donat això, s'ha de saber primer quins són els components que ha de tenir cada pàgina abans de programar-la per complir els requisits del projecte. El principal objectiu és fer consultes a la base de dades, per tant seran necessaris aquest components:

- Camps de text on escriure les dades que es volen consultar.
- Botó que indiqui que ja es pot fer la consulta sobre els camps indicats.
- Estructura que mostri els resultats d'aquests consultes.
- Estructura que permeti passar d'un tipus de consulta a una altra.

En conseqüència, es proposa el següent tipus de disseny per la web:

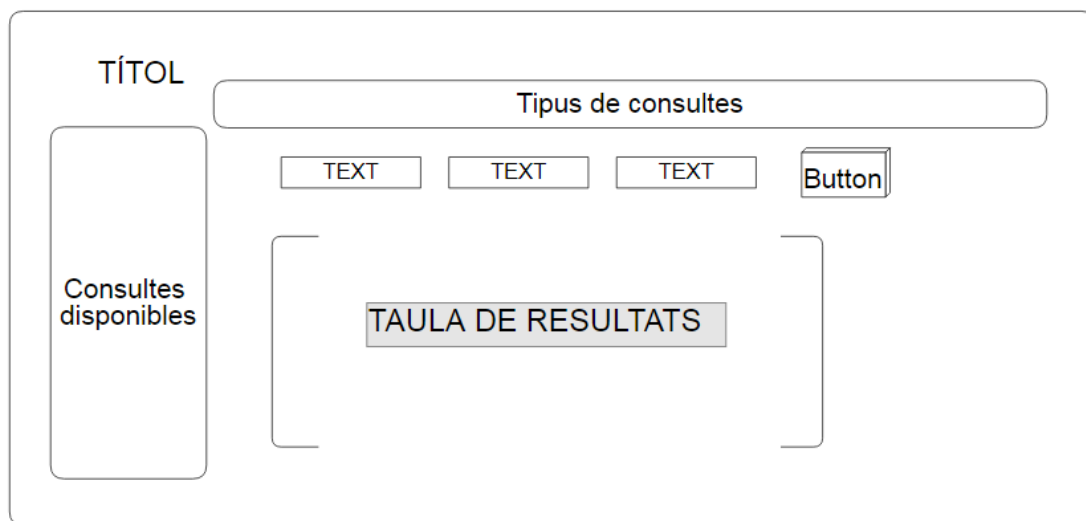


Figura 30. Disseny de la pàgina web

En el disseny web, la barra on s'especifiquen el tipus de consultes s'anomena "Navigation Bar" i on es mostren les consultes disponibles que aniran en funció de cada tipus s'anomena "Side Bar". [17]

Per crear els camps de text on s'escriuran les dades que es volen buscar, s'ha de crear un altre fitxer de *Python* anomenat `forms.py` (formularis) on es creen els objectes amb els camps necessaris per escriure les dades. Per exemple, un formulari adequat per les cerques del projecte seria aquest:

Nom: 1r Cognom: 2n Cognom: Data Inicial: Data Final:

Figura 31. Formulari per fer les consultes

En aquest formulari es defineixen 5 camps de text on es poden especificar el nom i els cognoms d'una persona i també les dates entre les quals es vol buscar.

Per mostrar els resultats de determinades consultes, s'han de crear taules i per això *Django* disposa d'un mòdul anomenat "`django_tables`". És necessària la creació d'un fitxer *Python* anomenat "`tables.py`" on es defineixin els tipus de taules adients per mostrar els resultats de les consultes. Per exemple, un tipus de taula on consultar les dades dels batejats seria aquest:

Arxiu	Nº	Data	Lloc	Nom	Noms Complementaris	Primer Cognom	Segon Cognom	Sexe	Data Naixement	Lloc Naixement	Observacions	Observacions Generals
B1(1596-1704)2	112	1651/10/19	—	Bonaventura	Francesch	Rodie	—	—	1651/10/19	Agullana	—	—

Figura 32. Exemple de taula pels resultats d'una consulta

Cada tipus de taula serà una classe del fitxer "`tables.py`" i cadascun dels seus paràmetres equivaldrà a una columna a la taula mostrada.

Ara bé, per tal de mostrar tots aquest components s'ha de dissenyar un fitxer HTML on es defineix el contingut i l'estructura de la pàgina web.

En conclusió, per crear cada pàgina de l'aplicació web caldrà pensar i seguir cadascun dels punts descrits en aquest ordre:

1. És necessària la creació d'un nou formulari per fer la consulta? O amb formularis anteriorment creats ja es compleixen els requisits d'aquesta?
2. És necessària la creació d'una nova taula per mostrar els resultats de la consulta? Les taules anteriorment creades es poden tornar a utilitzar per mostrar els resultats d'aquesta?
3. Quina URL es vol que tingui la pàgina? Afegir-la al llistat d'URL del fitxer "`urls.py`" i especificar quina funció ha de cridar del fitxer "`views.py`"
4. És necessària crear una nova funció o ja es pot reutilitzar una altre funció ja definida modificant-la per cobrir les necessitats de la consulta?.
5. Crear un fitxer HTML on es mostrin els resultats de la consulta.

Aquests serien els passos per crear cadascuna de les consultes del projecte que es descriuran en el següent apartat.

8.4. Consultes implementades

Per tal de organitzar el tipus de consultes possibles, es decideix dividir-les en quatre tipus.

- **General:** consultes que mostrin resultats sobres els tres tipus de registres que hi ha relacionant-los entre ells.
- **Baptismes:** consultes que només tractin els registres de baptismes.
- **Matrimonis:** consultes que només tractin els registres de matrimonis.
- **Obituàries:** consultes que només tractin els registres d'obituàries.

Donat això la pàgina inicial mostrarà tots els arxius dels quals es disposa amb els accessos als diferents tipus de registres i amb la possibilitat d'accedir a les consultes del tipus general:



Registres de la parroquia d'Agullana			
Arxius Disponibles			
Arxiu	Nom Parròquia	Persona Buidatge	Persona Revisat
B1(1596-1704)1	Agullana	RTF	RUM
B1(1596-1704)2	Agullana	RUM	RTF
B1(1596-1704)3	Agullana	JCP	PG
B1(1596-1704)4	Agullana	RTF	JOM
B1(1596-1704)5	Agullana	RTF	JOM
B2(1705-1744)1	Agullana	RUM	JOM
B2(1705-1744)2	Agullana	RUM	JOM
B2(1705-1744)3	Sta. Maria d'Agullana	ACS	JOM
B2(1705-1744)4	Sta. Maria A.	DV	FON
B2(1705-1744)5	Sta. Maria A.	DV	JOM

Figura 33. Captura de pantalla de la pàgina inicial

Cadascun dels tipus de consultes disposarà d'una cerca on es podran buscar els protagonistes dels esdeveniments en un rang d'anys determinat. L'usuari escriurà el nom, cognoms i el període d'anys entre els quals vol buscar i quan premi el botó "Cerca" obtindrà les columnes de la taula en la que està buscant.

Aquestes consultes consistiran en unes sentències SQL que demanaran totes les columnes de les taules i aniran afegint paràmetres a la clàusula del "where" si estan omplerts els camps de text dels formularis. Per exemple si es volen consultar els

baptismes de les persones anomenades “Pere” i de cognom “Clos” entre l’any 1800 i 1850, s’executarà la següent consulta a la base de dades:

```
select *  
from Baptisme  
where data_ini >= 1800 and data_fi <= 1850 and  
nom LIKE “%Pere%” and cognom1 LIKE “%Clos%”
```

Figura 34. Exemple consulta SQL

No serà necessari omplir tots els camps del formulari per fer efectiva la cerca. Per exemple, si no es desitja especificar l’any inicial o el final, doncs es cerquen tots els que compleixin les altres restriccions i si no s’especifica cap, doncs es mostren tots els resultats. A l’exemple anterior, no s’ha especificat el segon cognom i per això no s’ha afegit a les condicions.

També es pot veure a l’exemple anterior que en les condicions on apareixen els noms i els cognoms no s’utilitza el símbol d’igualtat si no la condició SQL “LIKE”, que busca dins de la cadena de caràcters la subcadena de caràcters que s’especifica. Això s’ha aplicat per donar més llibertat al resultat de les cerques i fer més accessibles altres resultats que potser no sortien tant. Per exemple, es podria donar el cas que el nom o cognom estigués al registre entre parèntesis, causant que una cerca per igualtat no considerés com a resultat les persones amb cognom “(Clos)” quan es busca “Clos”. Tot i això, sí que és restrictiu en el sentit de que busca exactament la mateixa cadena de caràcters, és a dir, si es busquen les persones amb cognom “clos” no sortien com a resultat les persones amb “Clos” ni “(Clos)”, per la diferència de la “C” majúscula.

Els resultats obtinguts de la consulta són tractats i transformats en la taula que es mostrarà a l’usuari.

Consultes generals

A l’apartat general, es podrà fer una cerca dels protagonistes en tots tres àmbits a la vegada, baptismes, matrimonis i obituaris. D’aquesta manera es pot intentar identificar quan va néixer una persona, quan es va casar i quan va morir. Aquesta cerca simplement consisteix en agrupar les tres cerques anteriors en una mateixa pàgina.

Una altra consulta de l’apartat general és la cerca de participacions d’una persona en determinats esdeveniments i el rol que ha exercit en aquests. Aquest cop la consulta es faria sobre la taula “participants” que conté tota l’informació dels participants en esdeveniments.

L'última consulta implementada a l'apartat general és l'anomenada "Vida". Aquesta consulta permet buscar, per una persona, tots els esdeveniments on ha estat protagonista i participant a la vegada. D'aquesta manera es podrien obtenir com petites biografies de les persones a partir del buidatge de registres.

Un exemple de resultat d'aquesta consulta podria ser el següent:

Registres de la parròquia d'Agullana

Consultes	General	Baptismes	Matrimonis	Obituaris
Protagonisme a esdeveniments Participacions a esdeveniments Vida	Per trobar els esdeveniments durant la vida d'una persona, introdueix les següents dades: Nom: <input type="text" value="Lorenzo"/> 1r Cognom: <input type="text" value="Suñer"/> 2n Cognom: <input type="text" value="Sallellas"/> Data Inicial: <input type="text" value="-"/> Data Final: <input type="text" value="-"/> <input type="button" value="Cerca"/>			
	Lorenzo Suñer Sallellas <ul style="list-style-type: none"> • Va néixer el dia 1835/07/15 a Agullana • Va ser batejat/ada el dia 1835/07/17 a Agullana • Va participar a un Baptisme com a Pare a data 1861/07/06 . • Va participar a un Baptisme com a Pare a data 1863/08/07 . • Va participar a un Baptisme com a Pare a data 1866/03/24 . 			

Figura 35. Captura de pantalla del resultat d'una consulta de Vida

En aquest exemple, s'ha buscat pels esdeveniments on ha participat la persona "Lorenzo Suñer Sallellas", on es pot veure quan va néixer, quan el van batejar i les seves participacions en baptismes, en aquest cas com a Pare i per tant sabem que en principi va tenir 3 fills. No figura el matrimoni ni l'obituari, el que vol dir que o bé no està registrat en aquestes dades perquè potser es va casar en una altra parròquia o bé perquè no complia determinats requisits que dicten que una persona en un registre és la mateixa en altres registres. El procés i condicions establerts per generar aquesta petita biografia de les persones són els següents:

1. Primer de tot, s'obtenen tots els baptismes en funció de les dates i els noms i cognoms especificats. D'aquesta manera ja es tenen identificats individus diferents.
2. Després s'obtenen els matrimonis amb els mateixos paràmetres que s'han utilitzat per buscar els baptismes. Per a cadascun dels individus identificats als baptismes, es busca el seu possible matrimoni. Per determinar-ho, s'estableix un sistema de puntuacions en el que es vagin sumant punts en funció de coincidències que puguin indicar que és el mateix individu. Si el noms coincideixen, es sumen 2 punts, si un dels noms és nul, es sumen 0.5 punts i s'aplica el mateix mètode per noms i cognoms. Si les dates d'inscripció o de naixement del baptisme tenen una diferència més petita de 15 o més gran de 70 amb la d'inscripció del matrimoni, la puntuació s'anul·la. Si al final l'individu identificat al matrimoni té una puntuació més gran o igual a 4,5 és considerat com a potencial mateixa persona i s'afegeix el matrimoni a la seva llista d'esdeveniments. Com es pot veure, no es considera que una persona es pugui casar abans dels 15 anys o després dels 70.

3. Seguidament, es procedeix a obtenir les participacions en diferents esdeveniments on no ha estat protagonista, també amb els mateixos paràmetres que s'han utilitzat per buscar els baptismes. Amb el mateix sistema de punts per els noms i cognoms que al pas anterior, aquest cop no es considera una participació si la data de l'esdeveniment va ser abans de 12 anys de la data de naixement o 65 anys després d'aquesta.
4. Per últim, es busquen el obituaries amb els mateixos paràmetres utilitzats pels baptismes i s'utilitza els mateix sistema de punts per noms i cognoms que al pas anterior. En els obituaries es considera que un individu és el mateix si el seu obituari es va registrar després de la seva data de naixement i 90 anys abans d'aquesta.

Com es pot veure, es bastant probable que per un mateix individu s'identifiquin els esdeveniments on va ser protagonista o va participar, però també hi ha la probabilitat de que se li atribueixin d'altres que potser no ho havia estat. Amb aquests errors s'haurà de conviure a causa de la natura de les dades.

Consultes de Baptismes

A l'apartat de baptismes, a més de les cerques generals també s'ha considerat interessant la cerca de vegades que una persona ha estat padrí o padrina i de qui. Per això s'ha implementat la consulta "Padrins i Fillols" on es poden fer cerques com la següent:

General

Baptismes

Matrimonis

Obituaris

Cerca el padrí/na introduint les seves dades:

Nom:

Pere

1r Cognom:

Clos

2n Cognom:

-

Data Inicial:

1770

Data Final:

-

Cerca

Participacions com a Padrí/na

Arxiu	Nº	Tipus Esdeveniment	Data Event	Tipus Relació	Nom	Primer Cognom	Segon Cognom	Estat Vital	Ofici	Lloc Naixement	Residència
B3(1745-1776)2	507	Baptisme	1770/04/06	Padri	Pere	Clos	—	viu/va	taper	—	Agullana
B4(1777-1805)	224	Baptisme	1799/10/29	Padri	Pere	Clos	—	viu/va	taper	—	—
B5(1805-1826)2	72	Baptisme	1813/09/30	Padri	Pere	Clos	—	viu/va	taper	—	—

Baptismes

Arxiu	Nº	Data	Lloc	Nom	Noms Complementaris	Primer Cognom	Segon Cognom	Sexe	Data Naixement	Lloc Naixement	Observacions	Observacions Generals
B3(1745-1776)2	507	1770/04/06	—	Llucia	Ursula, Justa	Pujarniscle	Puigagut	F	1770/04/05	Agullana	—	—
B4(1777-1805)	224	1799/10/29	—	Pere	Joseph, Joan	Clos	Carreras	M	1799/10/28	—	—	—
B5(1805-1826)2	72	1813/09/30	—	Pere	Joan, Joseph	Clos	Carreras	M	1813/09/29	—	—	—

Figura 36. Captura de pantalla del resultat d'una consulta de "Padrins i Fillols"

Consultes de Matrimonis

A l'apartat de matrimonis, a més de les cerques generals i al igual que amb els baptismes, també s'ha considerat interessant implementar una altra consulta que indiqui les participacions que ha tingut una persona com a Pare o Mare en els matrimonis, és a dir, la quantitat de fills que han sobreviscut als primers anys de vida i s'han casat. Un exemple de cerca podria ser el següent:

General

Baptismes

Matrimonis

Obituaris

Arxius de matrimonis disponibles. Cerca el matrimoni introduint les dades d'un dels pares:

Nom: 1r Cognom: 2n Cognom: Data Inicial: Data Final:

[Participacions com a Pare/Mare](#)

Arxiu	Nº	Tipus Esdeveniment	Data Event	Tipus Relació	Nom	Primer Cognom	Segon Cognom	Estat Vital	Ofici	Lloc Naixement	Residència
M2(1757-1805)2	206	Matrimoni	1798/02/03	Pare Marit	Pere	Clos	—	viu/va	taper	—	Agullana

[Matrimonis](#)

Arxiu	Nº	Data Inscripció	Data	Lloc	Nom Marit	Primer Cognom Marit	Segon Cognom Marit	Edat Marit	Lloc Naixement Marit	Estat Civil Marit	Residència Marit	Ocupació Marit	Nom Muller	Primer Cognom Muller	Segon Cognom Muller	Àlies Muller
M2(1757-1805)2	206	1798/02/03	1798/02/03	—	Joseph	Clos	Pujarnisda	—	—	S	Agullana	—	Elizabet	Carreras	Tarinas	—

Figura 37. Captura de pantalla del resultat d'una consulta de "Fills casats"

Consultes d'Obituaris

Pels obituaris, s'ha considerat interessant afegir una pàgina on es puguin veure les estadístiques de mortalitat infantil que es poden obtenir de les dades. Per fer això s'ha dividit en dècades el període que va des de 1707 fins a 1851 (el període de temps que assoleixen les dades dels obituaris) i s'han anat sumant els possibles candidats a ser nadons morts en el seu primer any de vida. Considerant que el camp on s'especifica l'edat està en format de text, és a dir, una cadena de caràcters, els passos i condicions que s'han establert per fer la classificació són aquestes:

- Es consideren tots els dígitos numèrics i s'obté la possible edat numèrica.
- Si el camp "edat" conté la lletra 'a' i no perquè estigui en 'dia', significa que l'individu no és un cas de mortalitat infantil perquè normalment quan està present la lletra 'a' en aquest camp és perquè forma part o és una abreviatura de 'anys', el que implica que ja tenia 1 any o més.
- Si el camp edat conté la lletra 'm' i l'edat numèrica calculada és més petita de 12, implica que hi havia la paraula 'mes' o 'mesos' i que en concret el nadó tenia menys de 12 mesos, per tant es considera mortalitat infantil.

- Si el camp edat conté la lletra 'd' referent a 'dia', o 's' referint-se a 'setmanes' o directament conté les paraules 'néixer' o 'neixer' o l'edat numèrica és igual a zero, llavors sí es considera mortalitat infantil.
- Si no es compleixen cap de les condicions anteriors llavors no es considera mortalitat infantil.

La gràfica que resulta després d'aplicar aquest mètode és la següent:

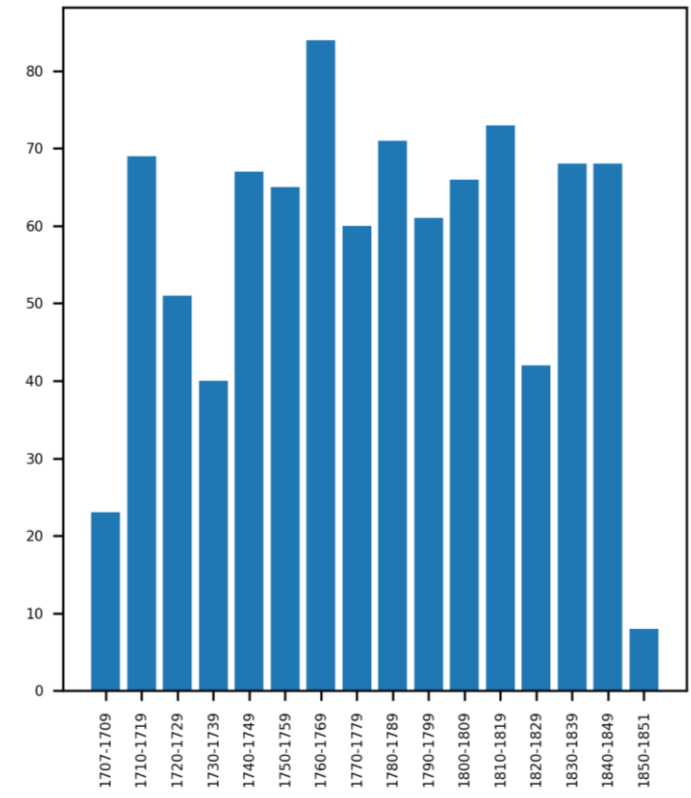


Figura 38. Gràfic on es representa la mortalitat infantil

Com es pot veure a la gràfica, la mortalitat infantil es manté pràcticament constant al llarg del període d'anys que es disposa, amb un pic a la dècada del 1760-1769 amb més de 80 obituaris de nadons registrats.

Detalls de les consultes

Un cop obtingut el resultat d'una consulta, s'obté un resultat limitat en comparació a com es trobaven abans als fitxers Excel, on a la mateixa fila es podien veure els protagonistes i els participants. El resultat de les consultes mostra exactament el que vol la consulta, per exemple quan es busca un batejat només es mostra informació sobre el batejat en sí, no sobre els participants que van formar part de l'esdeveniment, i el mateix passa del revés, quan es busca els participants no es pot saber qui va ser el protagonista, a excepció de les consultes "Padrins i fillols" i "Fills casats" però d'una manera molt limitada.

Per aquest fet seria interessant que, al prémer el botó esquerra del ratolí en cada línia de resultat d'una consulta, ens portés a un altra pàgina on es mostrés tota l'informació del registre de la qual ja es disposava als fitxers Excel.

Per realitzar aquesta funcionalitat, es va trobar el problema que *Django* no donava suport per crear-la, almenys des de les taules que donen el resultat d'una consulta. La solució a aquest problema va ser utilitzar Javascript [17] que, gràcies a la seva gestió d'esdeveniments, es pot generar un script que s'activés al prémer el botó esquerra en una fila de la taula, agafés el nom del llibre i el número de registre de la fila (clau primària composta de les taules) i se'ls enviés com a paràmetre a una funció de "views.py" que busqui tots els participants i protagonistes del registre en qüestió amb els paràmetres rebuts i els mostrés en una taula. Per exemple, una pàgina de detall podria ser aquesta al clicar sobre d'un resultat que referenciava un registre de baptisme:

General		Baptismes		Matrimonis		Obituaris							
Protagonista/es Esdeveniment													
Arxiu	Nº	Data	LLoc	Nom	Noms Complementaris	Primer Cognom	Segon Cognom	Sexe	Data Naixement	Lloc Naixement	Observacions	Observacions Generals	
B1(1596-1704)2	638	1670/04/15	—	Pere	Josep	Clos	—	—	1670/04/15	Agullana	—	—	
Participants Esdeveniment													
Arxiu	Nº	Tipus Esdeveniment	Data Event	Tipus Relació	Nom	Primer Cognom	Segon Cognom	Estat Vital	Ofici	Lloc Naixement	Residència		
B1(1596-1704)2	638	Baptisme	1670/04/15	Oficiant	Mateu	Cairol	—	viu/va	vicari i prevere	—	—		
B1(1596-1704)2	638	Baptisme	1670/04/15	Pare	Joan Antoni	Clos	—	viu/va	pagès	—	Agullana		
B1(1596-1704)2	638	Baptisme	1670/04/15	Mare	Anna	—	—	viu/va	—	—	—		
B1(1596-1704)2	638	Baptisme	1670/04/15	Padri	Joseph	Celler	—	viu/va	pagès	—	Ceret		
B1(1596-1704)2	638	Baptisme	1670/04/15	Padrina	Anna	Castelló	—	viu/va	—	—	Cantallops		
B1(1596-1704)2	638	Baptisme	1670/04/15	Conjuge Padrina	Pere	Castelló	—	viu/va	—	—	—		

Figura 39. Detall d'una consulta

Ara a partir de la cerca d'un batejat o de la qualsevol altre resultat de les consultes implementades anteriorment podem veure els participants de l'esdeveniment com, en cas de l'imatge anterior, l'oficiant, el pare, la mare, el padrí, la padrina ...

Vist que es disposa d'informació com dels progenitors o fins i tot dels avis del protagonista per determinades consultes, s'ha cregut adient la visualització d'un arbre genealògic que aportés valor al detall a partir de les dades disponibles.

Per crear aquest arbre genealògic, s'ha utilitzat el programa "graphviz" juntament amb el mòdul "pydot" de *Python*, que permeten la creació de grafs exportables a imatges.

Per consegüent, a partir de les dades que ens proporcionen els registres, podem obtenir per un registre de baptisme un graf d'aquest estil:

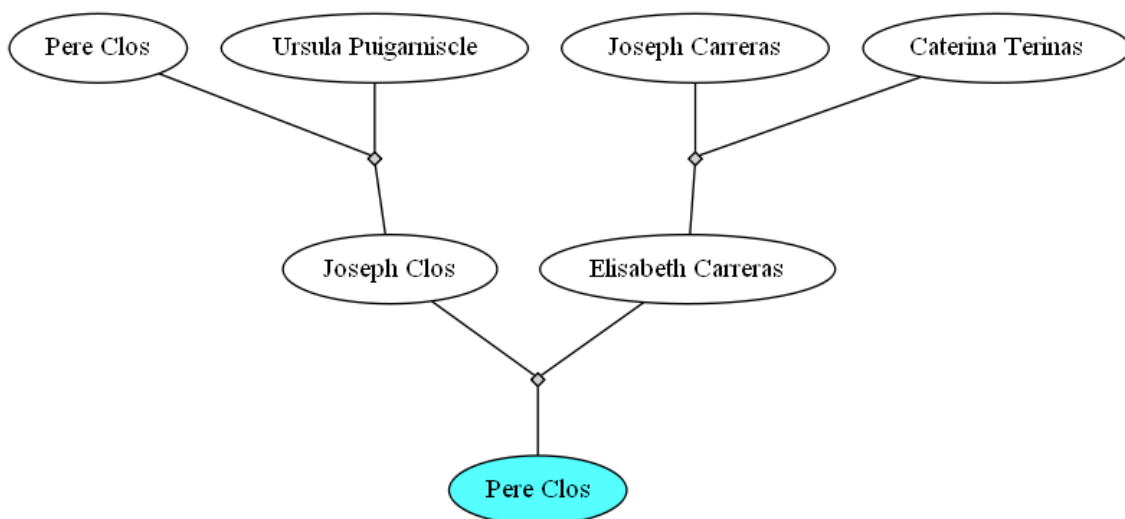


Figura 40. Arbre genealògic generat a partir de les dades del registre

Tot i això, amb les dades que es disposen es pot ampliar aquest arbre, per exemple es podrien buscar tiets, germans, parelles i fins tot fills dels protagonistes. Per fer efectiva aquesta funcionalitat, caldrà aplicar una sèrie de condicions similars a les aplicades a la consulta de “Vida” que ens permetin ampliar l’arbre genealògic de la manera més exacta possible.

Afegir germans del protagonista a l’arbre es pot fer en cas de disposar dels noms i cognoms dels pares. Simplement, cal fer una cerca de baptismes on els pares coincideixin amb els del protagonista. Si la data dels baptismes trobats es troba en un interval d’entre 15 anys abans i 15 anys després de la data de naixement del protagonista, es considera com a potencial germà i s’afegeix el nom i cognoms a les dades.

D’igual manera es poden afegir tant tiets paterns com materns, si es disposa dels noms dels avis respectivament. En aquest cas, es considera un tiet o tieta si aquest va néixer abans que el protagonista i fa menys de 40 anys.

A més, s’ha decidit buscar possibles fills del protagonista, així com les parelles amb els quals els va tenir. Es considera un possible fill del protagonista si aquest va néixer entre el període de 18 anys després i 50 anys abans del naixement del protagonista. Buscant els fills també es poden trobar les parelles del protagonista si estan especificades al registre. En funció de com estigui escrit, aquesta parella pot provocar que es creïn diferents nodes per a una mateixa parella. Per exemple, si es detecta un possible fill amb una parella de nom “Elisabet” i després es detecta un altre amb una parella de nom “Elisabert”, possiblement per un error quan es van escriure als fitxers, el programa ho detecta com dos parelles diferents i crea dos nodes per a cadascun quan en realitat només hauria de ser un node.

Però no només es poden generar arbres genealògics pels baptismes, també es pot pels altres tipus de registres que es disposen, matrimonis i obituaris.

En cas dels matrimonis es pot ampliar l'arbre genealògic buscant cunyats tant per part del marit com de la muller i també buscant fills de la parella recent casada. Es consideren cunyats si van néixer en un interval de 20 i 60 anys abans del matrimoni i fills si van néixer justament després del casament fins a 30 anys endavant.

Amb els obituaris també hi ha l'opció d'ampliar l'arbre. Si es disposa dels pares del difunt, es poden buscar els germans i si es disposa de la parella, també es poden buscar els fills. En cas dels germans, hauran d'haver nascut entre 10 i 60 anys abans de la defunció i en cas dels fills, entre abans de la defunció i 50 anys abans. En cas dels nadons que van morir al néixer o en els primers anys de vida no es detectarien els possibles germans que van néixer posteriorment.

Totes aquestes cerques es fan en els registres de baptismes, perquè es basen en buscar fills de determinades persones per ampliar l'arbre genealògic, encara que estiguin en qualsevol altre tipus de registre originalment.

Aplicant aquestes condicions es podran obtenir arbres genealògics d'aquest estil:

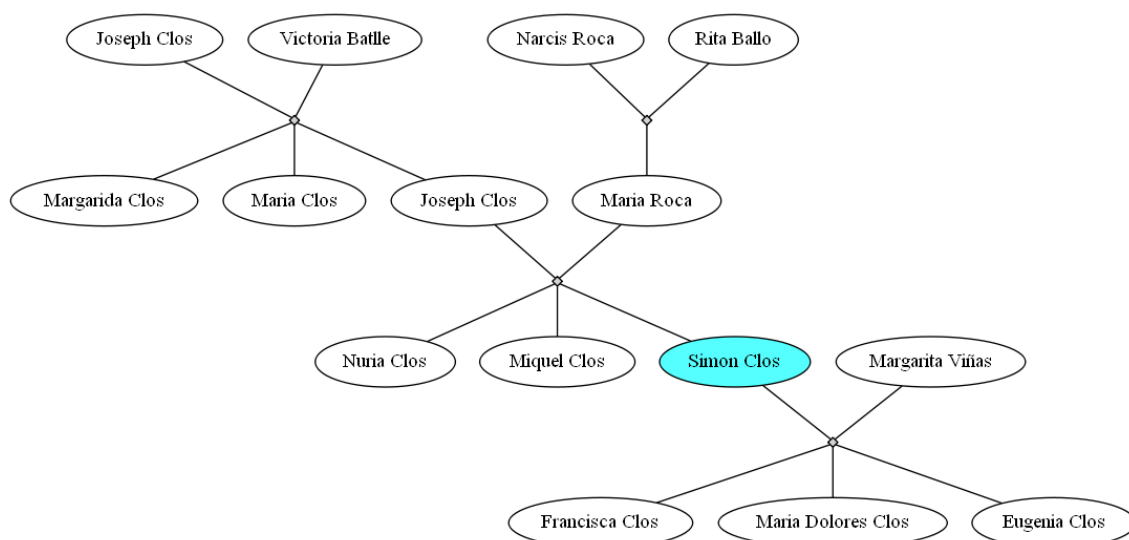


Figura 41. Arbre Genealògic generat a partir d'un registre de Baptisme

A la figura anterior, es marca en color blau el protagonista de la cerca que s'ha realitzat. En aquest cas, és el registre del baptisme de "Simon Clos". Gràcies a que es disposava de dades molt complertes en aquest registre, s'han pogut identificar als seus tiets per part de pare, als seus germans i fins i tot als seus fills i a la parella amb els quals els ha tingut. En definitiva, es poden apreciar fins a 4 generacions dels membres d'una família.

Tot el codi del projecte *Django* per l'aplicació web es pot consultar al corresponent enllaç a l'Annex.

9. Resultats

9.1. Objectius assolits

Amb la finalització d'aquest projecte, s'han aconseguit els seus dos objectius principals esmentats als apartats inicials. Primer, s'ha arribat amb èxit a migrar pràcticament tota la informació dels fitxers Excel a una base de dades relacional on hi ha més potencial per fer diferents tipus de consultes. Segon, s'ha implementat l'aplicació web connectada a la base de dades, des d'on es poden veure estadístiques com la mortalitat infantil, i fer anàlisis d'interès general, com per exemple consultar els familiars d'un batejat i obtenir el seu arbre genealògic.

En general, s'ha creat una eina que permet donar resposta a més preguntes de manera més completa. Amb aquest projecte ara es poden consultar avantpassats d'una manera més ràpida i còmode i fins i tot generar més informació a partir de les mateixes dades.

9.2. Futur del projecte

Aquest treball no es pot donar per tancat de forma definitiva. Encara hi ha la possibilitat de fer pública l'aplicació web, escollint un domini i contractant un servidor on allotjar-la. En cas de realitzar-se, no es preveu que tingui moltes visites degut a la poca varietat de dades de la que disposa i només interessaria a la gent de la població d'Agullana i als genealogistes de la zona. Tot i això, fer-la pública també significaria poder rebre feedback dels genealogistes amb experiència que podrien proposar més tipus de consultes interessants per ser implementades i augmentar l'interès i el valor de l'aplicació.

Ara bé, aquest projecte demostra que es pot fer una anàlisi interessant d'aquest tipus de dades, per exemple, associacions com la Societat Catalana de Genealogia podrien promoure la creació de projectes similars amb moltes més dades i per a un públic més ampli. Es podria reaprofitar el que ja està fet en aquest projecte per realitzar-ne d'altres de les mateixes característiques fins a un cert límit, ja que no s'ha pensat des d'un principi en fer-lo adaptable per d'altres, i a més tampoc es disposava d'altres dades. Per aquestes raons, només es podria ampliar aquest projecte amb més dades que tinguessin un format molt similar a les utilitzades en aquest projecte.

9.3. Aportacions del projecte

Per tal de portar a terme la realització d'aquest projecte s'ha hagut de passar per un procés d'aprenentatge necessari per aconseguir els objectius. Primer de tot, s'han adquirit els mètodes adequats per gestionar i preparar un projecte abans de començar a implementar-lo a la fase de Gestió i Economia de Projectes. Segon, s'han ampliat els coneixements que ja es tenien sobre el llenguatge de programació *Python*, cercant informació sobre com manipular arxius Excel i fer les connexions amb una base de dades. S'ha après a crear pàgines web, utilitzant eines àmpliament utilitzades com per exemple, el framework *Django*. En conclusió, l'execució d'aquest projecte ha contribuït al meu creixement i m'ajudarà en el desenvolupament de la meua carrera professional per haver tingut la oportunitat d'aplicar els coneixements i mètodes de treball adquirits durant la realització del grau d'enginyeria informàtica.

10. Annexos

Enllaç amb els scripts utilitzats per crear la base de dades i amb els scripts de *Python* utilitzats per fer el procés ETL:

<https://github.com/Nico152/ETL-Process-TFG>

Enllaç on es troba el codi de l'aplicació Web creada amb Django:

<https://github.com/Nico152/Analisi-de-registres-TFG>

11. Bibliografia

- [1]
[En línia]. Disponible a: <http://dag.cvc.uab.es/infoesposalles/>.
- [2]
«Ajuntament d'Agullana». [En línia]. Disponible a: <http://ca.agullana.cat/>.
- [3]
«Five centuries of marriages». [En línia]. Disponible a:
<http://dag.cvc.uab.es/projects/five-centuries-of-marriages/>.
- [4]
«Metodologies de Projectes» Treball de Fi de Grau de la UOC realitzat per Guifré Navarro i Nogués Disponible a:
<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/53504/7/gnavarroTFG0716mem%C3%B2ria.pdf>
- [5]
«Projectes». [En línia]. Disponible a:
<http://www.soci.scgenealogia.org/buidatges/projectes.htm>.
- [6]
«Registros gratuitos de historia familiar y genealogía — FamilySearch.org». [En línia].
Disponible a: <https://www.familysearch.org/>.
- [7]
«Societat Catalana de Genealogia ». [En línia]. Disponible a:
<https://www.scgenealogia.org/>.
- [8]
«Estudi de Remuneració 2017 en Tecnologia de Michael Page» Disponible a:
https://www.michaelpage.es/sites/michaelpage.es/files/MP_SPA_ON_ER_IT_03052017.pdf
- [9]
«Cacoo». [En línia]. Disponible a: <https://cacoo.com/es/>.
- [10]
«Requirements Engineering: A Roadmap». Nuseibeh, B.; Easterbrook S.; ICSE'2000,
IEEE Computer Society Press
- [11]
«PostgreSQL – Full Stack *Python*». [En línia] Disponible a:
<https://www.fullstackpython.com/postgresql.html>
- [12]
«Software Engineering. A Practitioner's Approach». Pressman, R.G.; Mc Graw-Hill,
2005 (7ª Edició)
- [13]
«Database Systems Implementation». Garcia-Molina, H.; Ullman, J.; Widom, J.;
Prentice-Hall, 2000.
- [14]
«Python Data Analysis Library». [En línia] Disponible a: <https://pandas.pydata.org/>

- [15]
«Psycopg – PostgreSQL database adapter for Python». [En línea] Disponible a:
<http://initd.org/psycopg/docs/>
- [16]
« ¿Qué es Django? – Django Girls Tutorial». [En línea] Disponible a:
<https://tutorial.djangogirls.org/es/django/>
- [17]
«W3Schools Online Web Tutorials». [En línea] Disponible a:
<https://www.w3schools.com/>
- [18]
«The Web Framework for perfectionist with deadlines - Django». [En línea] Disponible
a: <https://www.djangoproject.com/>